STRUCTURAL AND INHIBITION STUDIES ON UDP-GALACTOPYRANOSE MUTASE

A Thesis Submitted to the College of Graduate Studies and Research

in Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

In the Department of Chemistry

University of Saskatchewan

By

Sarathy Karunan Partha

Permission to Use

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Chemistry
University of Saskatchewan
Saskatoon, Saskatchewan (S7N 5C9)

ABSTRACT

UDP-galactopyranose mutase (UGM) is a flavoenzyme which catalyzes the interconversion of UDP-galactopyranose (UDP-Gal*p*) and UDP-galactofuranose (UDP-Gal*f*). UDP-Gal*f* is the active precursor of Gal*f* residues. Glycoconjugates of Gal*f* residues are found in the cell wall of bacteria and on the cell surface of higher eukaryotes. Gal*f* residues have not been found in humans and the fact that they are essential for the growth of pathogenic bacteria makes UGM a potential antibacterial target.

In the present study, crystal structures of UGM from *Deinococcoccus radiodurans* (drUGM) in complex with substrate (UDP-Gal*p*) were determined. UDP-Gal*p* is buried in the active site and bound in a U-shaped conformation. The binding mode and active site interactions of UDP-Gal*p* are consistent with the previous biochemical and mechanistic studies. The mobile loops in the substrate complex structures exist in a closed conformation and Arg198 on one of the mobile loops stabilizes the phosphate groups of the substrate. The anomeric carbon of galactose is 2.8 Å from the N5 of FAD (in the reduced complex) favorable to form FAD-galactosyl adduct. In addition to substrate complex structures, the crystal structures of drUGM in complex with UDP, UMP, and UDP-Glc have been determined. The mobile loops in all these complexes exist in a closed conformation.

Inhibitors for UGM were identified by ligand-based and structure-based methods. The phosphonate analog of UDP-Gal*p* (GCP) showed only weak inhibition against various bacterial UGMs. The structure of drUGM in complex with GCP provided a basis for its inhibitory activity. Poor stabilization of the phosphate groups by conserved arginines (Arg198 and Arg305) and altered sugar binding mode account for its activity.

Novel indole-based (LQ1, LQ6 and LQ10) inhibitors of UGM were identified through structure-based virtual screening (SBVS) of a chemical library. Inhibition studies also allowed the identification of an active site aspartic acid that plays role in inhibitor binding.

The structural studies on drUGM provided a basis for understanding substrate binding to UGM. *In vitro* enzyme inhibition studies allowed the identification of novel indole-based inhibitors. The structural and inhibition studies reported here enhance the understanding of UGM-ligand interactions and will assist in the development of more potent inhibitors of UGM.

ACKNOWLEDGEMENTS

I feel deeply indebted to my supervisor, Dr. David Sanders, for giving me this opportunity of carrying out this exciting project in his laboratory. I am most thankful for his never-ending interest in the progress of this project and for teaching protein X-ray crystallography. The amount of knowledge and confidence I gained from him during these years is immense. I thank Dr. Sanders for showing huge interest to answer my questions seriously and for all his scientific and non-scientific advice throughout my PhD program. I thank Dr. Sanders, for providing me the opportunity to attend American Crystallographic Association 2009, Toronto to present some of the structural results of this dissertation.

I would like to thank Dr. David Palmer for his invaluable suggestions for the project and answering my questions without any hesitation. It was a great pleasure and learning experience with Dr. David Palmer and I enjoyed the classes he taught to me.

I would like to thank the present and past members of Dr. Sanders' lab for making my stay at UofS enjoyable and memorable. I thank Dr. van Straaten for spending a lot of time in teaching me X-ray crystallographic techniques from data collection to structure solution. Special thanks to Josiah for helping me to learn many basic lab techniques in the initial stages of my program.

I would like to thank the committee members, Dr. Grochulski, Dr. Pedras and Dr. Reid for their constructive criticism and valuable suggestions. I would like to thank Department of Chemistry for providing me the financial support over the period of my

studies. Also, I would like to thank all the staff members (Virginia, Kathy, Dwight, Tanis, Devin, Garth, Ronda), Department of Chemistry for all their help and support.

I would like to thank Dr. David Jakeman (Dalhousie University) for generously providing samples of UDP-Gal*p* analog for structural and inhibition studies. I express my sincere thanks to Dr. Todd Lowary (University of Alberta) for accepting me to visit his lab to gain technical expertise (UDP-Gal*f* synthesis).

I would like to thank my parents Karunan and Sarojini who generously allowed me to pursue doctoral studies abroad. Their blessings and support is a "vitamin" for me to achieve this task. I am grateful to have my wife Viji for her tremendous love, support, motivation and understanding throughout this program. I would also like to thank my sisters (Geetha and Lakshmi) and brother (Bhaskar) for their support and encouragement.

Finally, I am grateful to have friends like Visu, Venkat and Anand from my undergraduate studies back in India. I am unable to put in words about their support and encouragement over the period of my study. I would like to thank all the friends and families in Saskatoon for their support throughout the period of my study. My special thanks to Dr. Swarnam Ravindran and Vijay Mamillapalle for their support throughout the good and bad times.

DEDICATION

I dedicate this dissertation to my
Parents
Sarojini and Karunan

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AEBSF	4-(2-aminoethyl)-benzenesulfonyl fluoride
asu	asymmetric unit
cAMP	cyclic adenosine monophosphate
CCP	collaborative computational project
cGMP	cyclic guanosine monophosphate
dCMP	deoxycytidine monophosphate
DHF	dihydrofolate
DHFR	dihydrofolate reductase
DNase	deoxyribonuclease
dNTP	deoxyribonucleotide triphosphate
dTMP	deoxythymidine monophosphate
dUMP	deoxyuridine monophosphate
FAD	flavin adenine dinucleotide
GALU	glucose-1-phosphate uridyltransferase
GalPUT	galactose-1-phosphate uridyltransferase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	high performance liquid chromatography
IDA	iminodiacetic acid
IPTG	isopropyl-β-thiogalactoside
LB	Luria-Bertani
LPS	lipopolysaccharides
LTA	lipoteichoic acid

MIC	minimum inhibitory concentration
MR	molecular replacement
NADH	nicotinamide adenine dinucleotide (reduced)
NADPH	nicotinamide adenine dinucleotide phosphate (reduced)
NTA	Nitrilotriacetic acid
PCR	polymerase chain reaction
PDB	protein data bank
PDE4D3	phosphodiesterase 4D3
PEG	polyethylene glycol
rmsd	root mean square deviation
rpm	rotations per minute
SBVS	structure-based virtual screening
SDS-PAGE	sodium-dodecyl sulfate polyacrylamide gel electrophoresis
SET	single electron transfer
STD-NMR	saturation transfer difference-NMR
TB	tuberculosis
THF	tetrahydrofolate
TS	thymidylate synthase
UDP	uridine diphosphate
UGM	UDP-galactopyranose mutase
UMP	uridine monophosphate
UTP	uridine triphosphate
WT	wild type

TABLE OF CONTENTS

CHAPTER 3 Materials and Methods

CHAPTER 4 drUGM-ligand structures

CHAPTER 5 Evaluation of UGM inhibitors

CHAPTER 1

Introduction

1.0 Bacterial cell wall as a drug target

The cell wall of bacteria is composed of proteins, lipids and carbohydrates and its architecture varies between Gram-positive and Gram-negative bacteria (Green, 2002). The cell wall of Gram-positive bacteria consists of a peptidoglycan layer and lipoteichoic acid (LTA). LTA extend through the peptidoglycan layer and appear on the cell surface (Figure 1-1a). In Gram-negative bacteria the cell wall is made of thin peptidoglycan layer and an outer cell membrane (Figure 1-1b) that contains lipopolysaccharides (LPS), phospholipids and proteins (Green, 2002). LPS O-antigen is an important virulence factor in Gram-negative bacteria and is required for resistance to complement-mediated serum killing (Joiner, 1988). The porins in Gram-negative bacteria regulate the passage of molecules across the outer membrane and are embedded within the lipid bilayer. In addition, Gram-negative bacteria also have a gel like periplasm between the cell wall and outer membrane. The cell wall of Gram-positive bacteria is thicker than the cell wall of Gram-negative bacteria. This is due to the presence of 20-fold more peoptidoglycan in Gram-positive bacteria. The cell wall of *Mycobacterium tuberculosis*, the causative organism of tuberculosis (TB), is a complex structure with an outer mycolic acid layer connected to the peptidoglycan through an arabinogalactan layer (Figure 1-2) (Pedersen & Turco, 2003; Weston et al, 1997).

The bacterial cell wall helps in maintaining osmotic pressure inside the cell and protects the cellular components from the external environment. Also, the cell wall defines the shape of the cell. The integrity of the cell wall is critical especially during

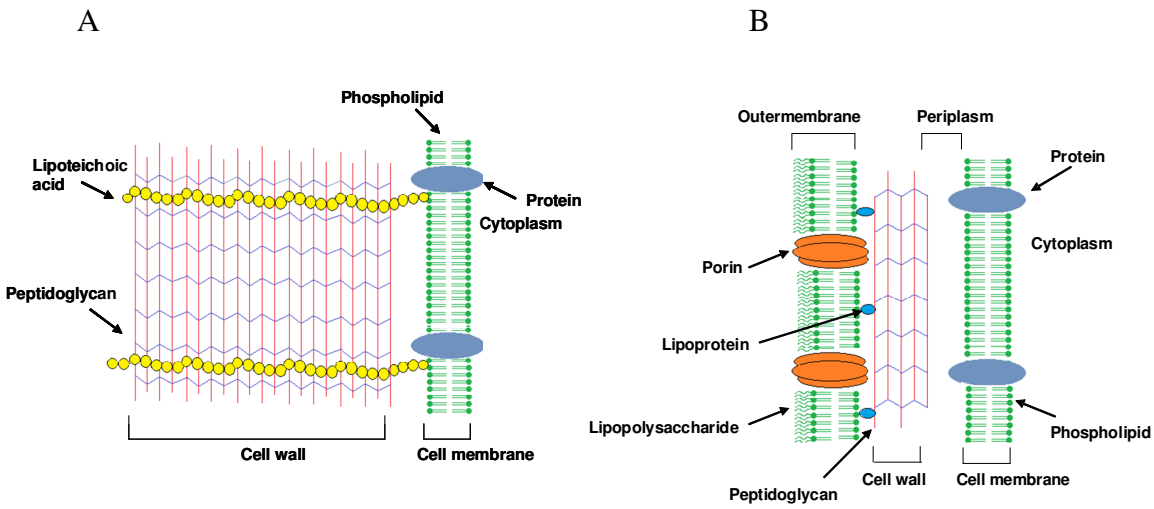


Figure 1-1 Cell wall architecture of Gram-positive (A) and Gram-negative bacteria (B).

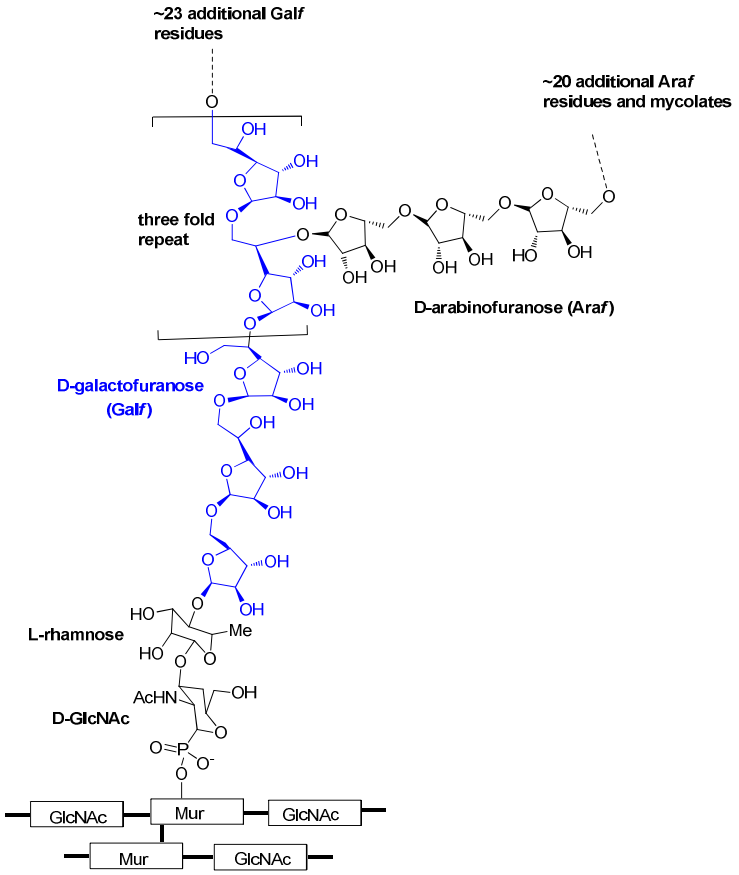


Figure 1-2 Schematic representation of the cell wall architecture of *Mycobacterium* to highlight the presence of Gal*f* residues (GlcNAc = N-acetylglucosamine, Mur = Muramic acid.

cellular replication and growth. Any deleterious effect to the cell wall architecture can alter the cell wall integrity resulting in leakage of cellular components and eventually cell lysis. Thus, the bacterial cell wall is a potential drug target for developing novel antibacterial agents. Antibacterial therapeutics like penicillins and cephalosporins act by interrupting the cell wall biosynthesis.

1.1 Biological significance of galactofuranose

Carbohydrates are integral components of the cell wall of bacteria. Biochemical characterization of the cell wall components of various pathogenic microorganisms revealed the presence of a unique glycoconjugates made of D-galactofuranose (Gal*f*) residues. For example, Gal*f* conjugates are found in the LPS O-antigen of a number of pathogenic Gram-negative bacteria, including *Klebsiella pneumoniae* (Köplin et al, 1997), *Escherichia coli* (Nassau et al, 1996), *Shigella dysenteriae, Shigella boydii, Serratia* spp., *Actinobacillus pleuropneumoniae* and *Pasteurella hemolytica* (Knirel & Kochetkov, 1994). In addition, Gal*f* residues are essential components of the arabinogalactan layer of the *Mycobacterium* cell wall (Figure 1-1) (Weston et al, 1997) and are essential for their growth (Pan et al, 2001). Gal*f* residues are also found in the cell wall and cell surface structures of other pathogenic micoroorganisms like *Aspergillus* (fungi) and *Trypanosoma* (protozoa) (de Lederkremer et al, 1980; Takayanagi et al, 1994). Development of resistant strains of pathogenic bacteria to current antibacterial agents complicates treatment (Green, 2002). Furthermore, the treatment of TB is complicated by patient compliance, side effects, and emergence of multi-drug resistant (MDR) strains of *M. tuberculosis* to current anti-TB agents (Kremer & Besra, 2002). Hence, there is an urgent need to identify novel antibacterial drug targets and drugs

acting on them (Zhang, 2005; Green, 2002). Gal*f* biosynthesis has been proposed as a potential target for developing novel antibacterial therapeutics (Pedersen & Turco, 2003).

1.2. Role of UGM: Identification and Characterization

Sugar residues like Gal*f* are incorporated into the cell wall of bacteria by using sugar nucleotides as donor molecules. The sugar nucleotide, UDP-Gal*f* acts as a active precursor or donor molecule for Gal*f* residues (Weston et al, 1997). UDP-Gal*f* is synthesized from UDP-Gal*p* by the enzyme UDP-galactopyranose mutase (UGM) (Weston et al, 1997). UGM catalyzes the reversible conversion of UDP-Gal*p* (pyranose ring) and UDP-Gal*f* (furanose ring) and the equilibrium favors the pyranose form over the furanose form (Figure 1-3) (Sanders et al, 2001; Zhang & Liu, 2000; Zhang & Liu 2001). The first step in the biosynthesis of UDP-Gal*f* (Figure 1-4) is the phosphorylation of Gal*p* to form galactose-1-phosphate. In a subsequent uridylation step a UMP unit is incorporated into galactose-1-phosphate to form UDP-Gal*p*. In the final step, UDP-Gal*p* is converted to UDP-Gal*f* by the action of UGM. Alternatively, UDP-Gal*p* can be synthesized from UDP-glucose by the action of UDP-galactose-4-epimerase (Thoden & Holden, 1998). Upon formation of UDP-Gal*f* various galactosyl transferases incorporate the Gal*f* residues into the cell wall by using UDP-Gal*f* as a sugar donor (Rose et al, 2006). The fact that UGM is essential for the growth of certain pathogenic bacteria and

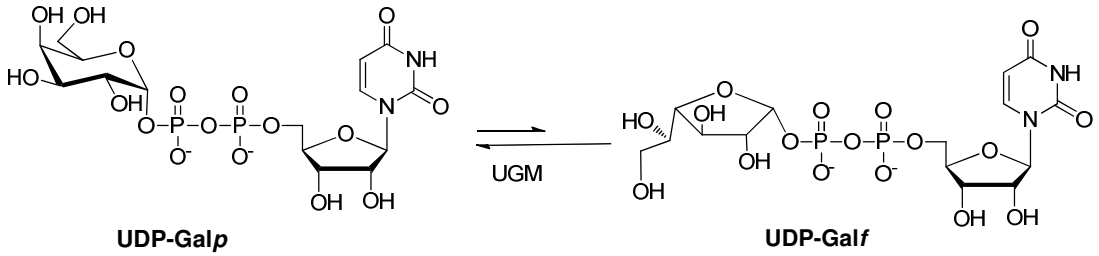


Figure 1-3 UGM catalyzed interconversion of UDP-Gal*p* and UDP-Gal*f*. Equilibrium favors the pyranose (UDP-Gal*p*) form over the furanose (UDP-Gal*f*) form.

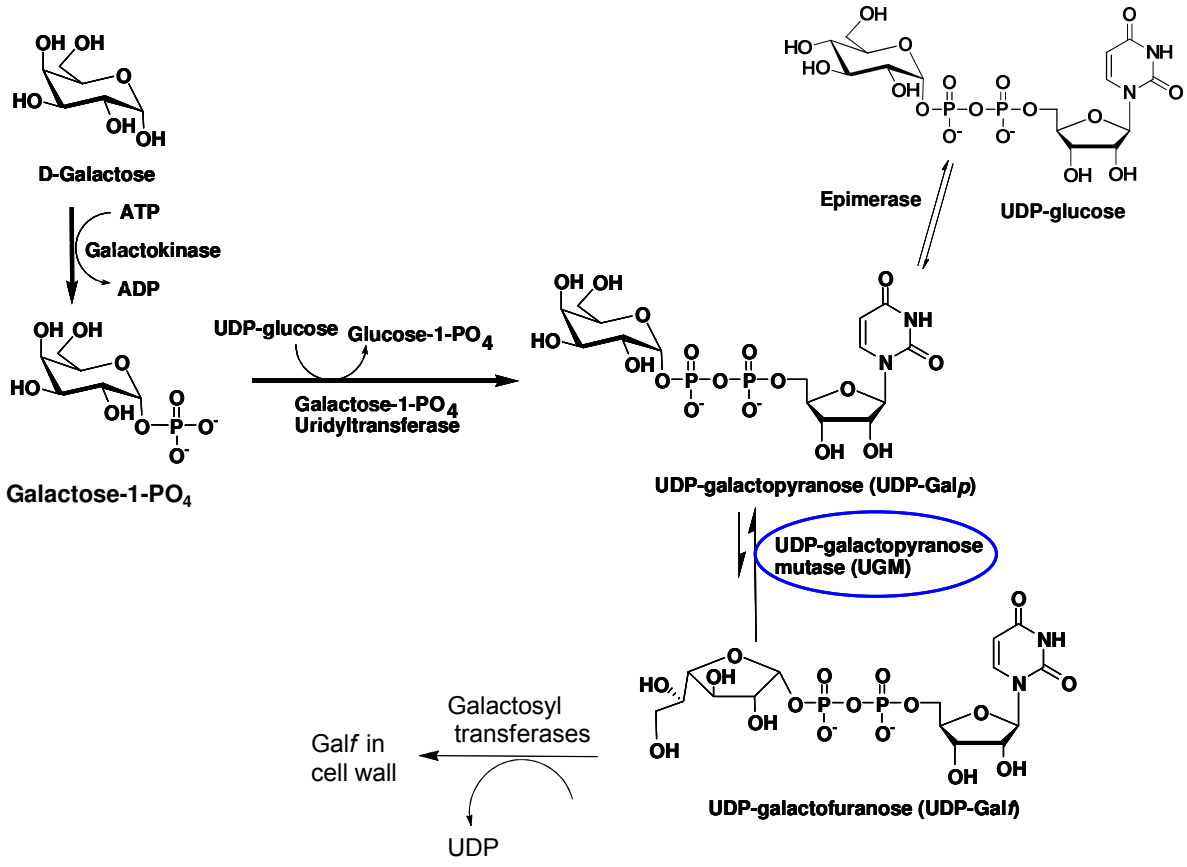


Figure 1-4 Biosynthesis of UDP-Gal*f*. UGM is found only in certain pathogenic microorganisms and have not been found in humans.

its absence in humans suggests that UGM is an ideal target for developing novel antibacterial therapeutics. The genes encoding UGM in *E. coli* (ecUGM), *K. pneumoniae* (kpUGM), and *M. tuberculosis* (mtUGM) have been identified, cloned, expressed and the enzymes have been purified and characterized (Köplin et al, 1997; Nassau et al, 1996; Weston et al, 1997). ecUGM is the first enzyme from any organism demonstrated to catalyze the isomerization of pyranose form of a sugar to furanose form (Köplin et al, 1997). UGM is a flavo-enzyme and contains the cofactor flavin adenine dinucleotide (FAD). FAD is bound to the enzyme non-covalently. The presence of FAD in UGM was shown by UV-visible spectroscopy with characteristic peaks at A_{382} and A_{450}, with a

shoulder at A_{465} unique to FAD (Nassau et al, 1996). The purified fractions of UGM were yellow in color due to the presence of FAD. SDS-PAGE analysis and mass spectroscopy of ecUGM revealed the approximate molecular mass of the enzyme (about 45 kDa). Subsequent to the identification of ecUGM, the gene that encodes kpUGM was cloned, over-expressed, purified and characterized (Köplin et al, 1997). The molecular weight of kpUGM was found to be 45 kDa (by SDS-PAGE), similar to the molecular weight ecUGM. However, gel filtration analysis of kpUGM revealed a molecular weight of 92 kDa, suggestive of a dimeric structure for the native enzyme (Köplin et al, 1997). The role of UGM in *Mycobacteria* was shown by gene-knockout experiment using *M. smegmatis* as a model organism (Weston et al, 1997) and it has been shown that Gal*f* biosynthesis is essential for growth. This suggests that mtUGM is a potential target for developing novel anti-TB drugs (Pan et al, 2001).

1.3. Biochemical investigations on UGM

Flavoenzymes are usually involved in the catalysis of redox reactions (Mansoorabadi et al, 2007). In case of UGM, the interconversion of UDP-Gal*p* and UDP-Gal*f* does not change the redox state of the substrate, i.e., there is no net gain or loss of electrons in the reaction (Sanders et al, 2001). UGM activity was not observed in the absence of reductant NADH or NADP(H), suggestive of the need for the enzyme to be reduced for activity (Köplin et al, 1997; Nassau et al, 1996). Based on the ecUGM crystal structure, it has been proposed UGM does not have a specific binding cavity or pocket available for NADH or NADPH. In addition, sodium dithionite was also found to be an effective reducing agent for FAD in UGM (Sanders et al, 2001). Interconversion of UDP-Gal*p* and UDP-Gal*f* was observed when sodium dithionite was used as a reducing agent

(Sanders et al, 2001). These preliminary investigations assign a specific role for FAD in the reaction mechanism. UGM activity experiments performed by adding the oxidant (K_3FeCN_6) to fully reduced UGM abolished its activity. These experiments clearly showed that oxidized UGM is inactive and only reduced enzyme is active (Sanders et al, 2001). These studies assign a specific role for FAD in the reaction and it must be reduced for activity. The role of N5 atom of the isoalloxazine ring of FAD in UGM was demonstrated by reconstitution of UGM with the FAD analogs (Figure 1-5), 1-deaza-FAD and 5-deaza-FAD (Huang et al, 2003). The activity of UGM reconstituted with 1-deaza-FAD was comparable to the wild type, but the 5-deaza-FAD analog did not show any activity. Therefore, it is evident that N5 of FAD is critical for UGM activity and participates in the reaction mechanism.

One of the proposed steps in the interconversion of UDP-Gal*p* and UDP-Gal*f* is the cleavage of the anomeric C-O bond (Barlow et al, 1999). Evidence for the cleavage of the anomeric C-O bond is supported by positional isotope experiments performed using labeled UDP-Gal*p* (^{18}O at anomeric position and ^{13}C at C1 position). The cleavage of C-O bond was monitored by detecting the isotopic shift accompanied by scrambling of ^{18}O from the bridging position into a non-bridging position (Barlow et al, 1999).

The mechanism of UGM reaction was further investigated using fluoro-substituted substrate analogs as probes and also characterization of reaction intermediates (Soltero-Higgin et al, 2004). Based on these studies, three different mechanisms, namely S_N1, S_N2 and single electron transfer (SET), have been proposed (Figure 1-5 and 1-6) for the UGM reaction (Fullerton et al, 2003; Soltero-Higgin et al, 2004; Zhang & Liu 2001). The first step in the S_N1 pathway involves the formation of an oxocarbenium ion by the

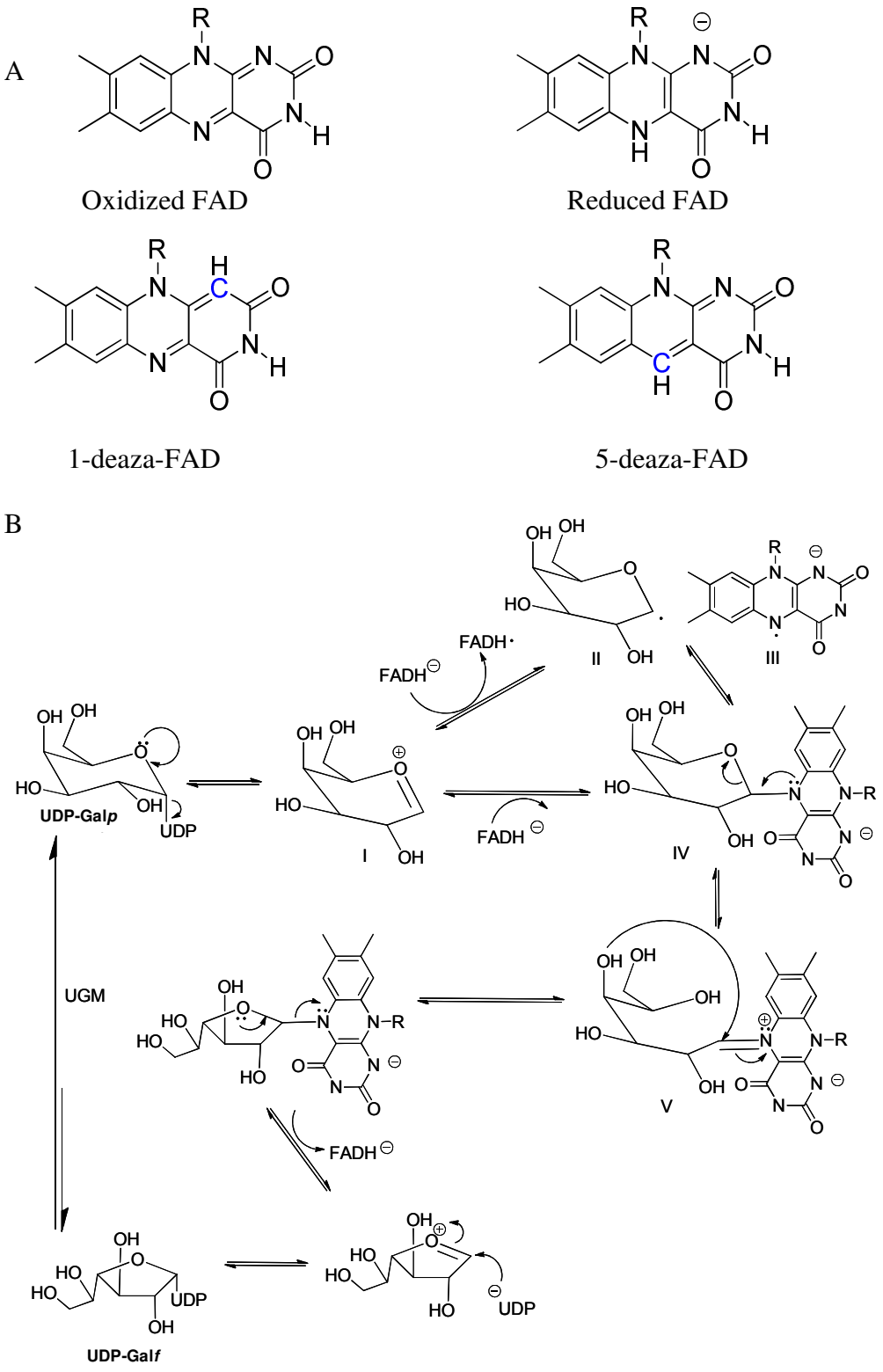


Figure 1-5 FAD structure and proposed S_N1 or SET pathways for UGM mediated interconversion of UDP-Gal*p*. **A)** Structures of oxidized/reduced FAD and deaza-FAD analogs. **B)** Proposed S_N1 or SET pathways for UGM mediated interconversion of UDP-Gal*p* and UDP-Gal*f*. I, oxocarbenium ion; II, galactosyl radical; III, semiquinone form of FAD; IV, FAD-galactosyl adduct; V, iminium ion intermediate.

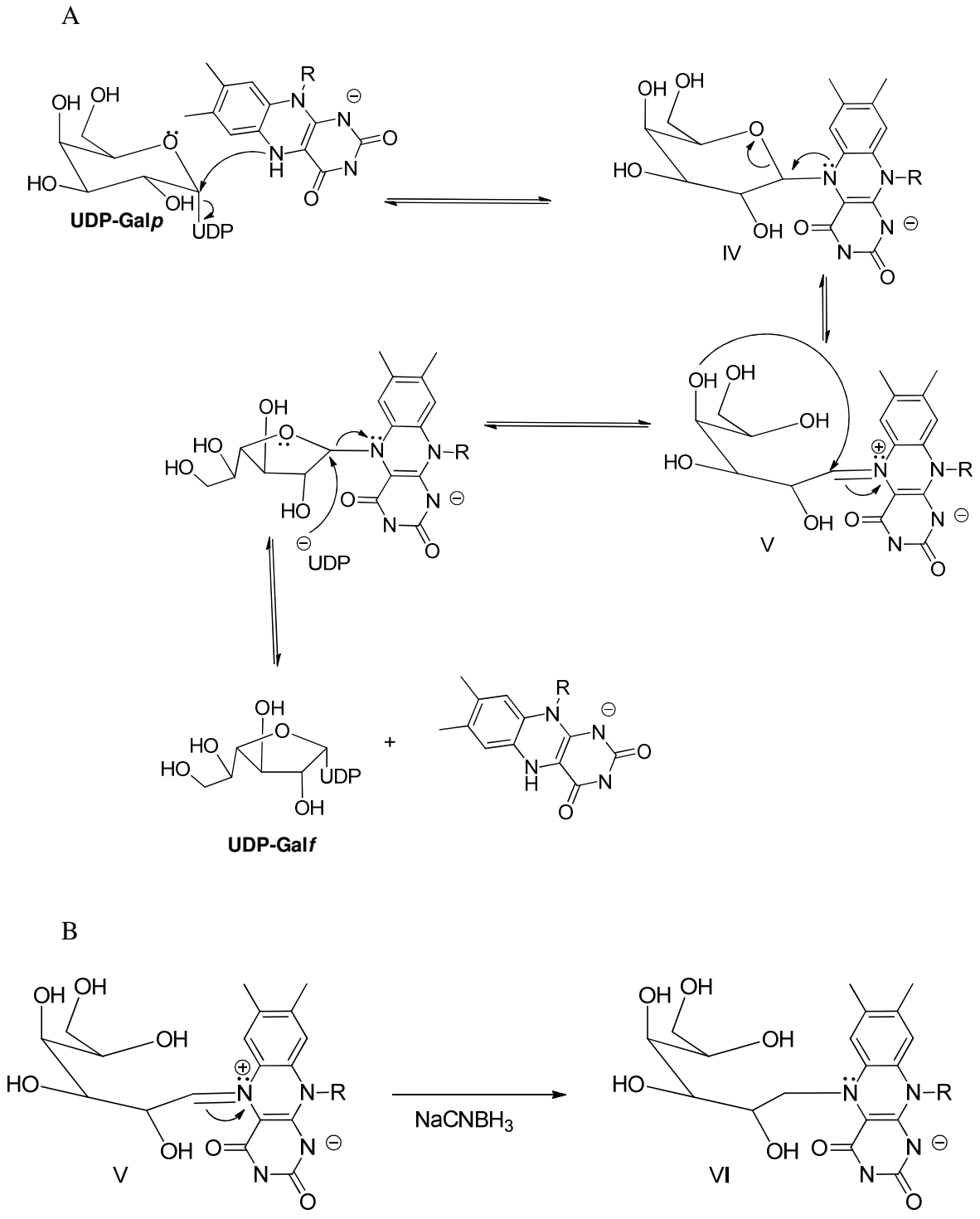


Figure 1-6 Proposed S_N2 pathway for UGM reaction. **A)** Concerted S_N2 mechanism for UGM reaction. The intermediates (for example IV and V) are similar to the S_N1 pathway. **B)** Sodium cyanoborohydride reduction of IV to form the adduct VI. Adduct VI is characterized by mass spectrometry.

cleavage of the anomeric C-O bond to release UDP. The reduced FAD acts as a nucleophile and attacks the C1 of the galactose (in oxocarbenium ion form) to form an FAD-galactosyl adduct. In a ring opening step, the FAD-galactosyl adduct collapses to form a iminium ion intermediate. Subsequently, the iminuim ion undergoes a ring closure to form an FAD-galactofuranose intermediate. In an S_N1 manner, UDP attacks the C1 of galactofuranose (in oxocarbenium ion form) to form UDP-Gal*f*. In support of S_N1 or oxocarbenium ion formation, various fluoro-substituted analogs of UDP-Gal*p* and UDP-Gal*f* have been synthesized and tested against UGM (Barlow & Blanchard, 2000; Zhang & Liu 2001). Presence of fluorine atom (2-fluoro and 3-fluoro substitution at the sugar moiety) would destabilize the oxocarbenium ion formation through inductive effect and may act as an inhibitor or poor substrates for UGM. Indeed, the fluoro-susbtituted analogs of UDP-Gal*p* and UDP-Gal*f* were found to be poor substrates for UGM, which indirectly supports the hypothesis that an oxocarbenium ion intermediate forms during the reaction. In a recent study, an open chain analog of UDP-Gal*p* (UDP-Gal-OH) has been designed, synthesized and tested as a substrate for UGM (Itoh et al, 2007). The open chain analog did not show any turnover, but showed inhibitory activity towards UGM and this is considered as indirect evidence for S_N1 pathway.

An S_N2 mechanism is also proposed for UGM reaction (Soltero-Higgin et al, 2004), which involves the direct attack of N5 of $FADH^-$ on the anomeric carbon galactose to form an FAD-galactosyl adduct with simultaneous release of UDP. The FAD-galactosyl adduct then undergoes ring opening and ring closure reactions to form FAD-galactofuranose. In the final step, UDP attacks the C1 of galactofuranose to form UDP-Gal*f* and regenerate the reduced FAD. The formation of the FAD-galactosyl adduct

has been demonstrated by reducing the iminium ion intermediate with sodium cyanoborohydride and the reduced intermediate characterized by mass spectrometry (Soltero-Higgin et al, 2004).

Alternatively, the FAD-galactosyl adduct has been proposed to form via SET from reduced FAD to the oxocarbenium ion to form a galactosyl radical and the semiquinone form of FAD (Fullerton et al, 2003). The two radical species then recombine to form an FAD-galactosyl adduct. Potentiometric analysis of UGM provides evidence for the stabilization of the semiquione form of FAD upon binding of the substrate and also suggests that the fully reduced flavin in kpUGM has anionic character ($FADH^-$) (Figure 1-5). The proposed SET pathway has been challenged on the basis of the redox potential of oxocarbenium ion and the FAD (Soltero-Higgin et al, 2004). The midpoint reduction potential of UGM was found to be -97 mV. Although the reduction potential of cylic galactose oxocarbeniuim ion is unknown, the midpoint potential of protonated formaldehyde was found to be -1,182 mV. For electron transfer to occur the potential of donor/acceptor must be close enough and the reduction potential values suggest that the reduced FAD in UGM may not be a strong enough reducing agent for the oxocarbeniuim ion to generate galactosyl radical.

In summary, three different mechanisms have been proposed for the UGM reaction and a major difference among them is the mechanism of formation of the FAD-galactosyl adduct. Evidence for and against the proposed mechanisms further complicates the interpretation of this unique enzymatic reaction. Further biochemical investigations are required to understand the precise mechanism of UGM reaction, which is beyond the scope of the current study.

1.4 Structure of UGM

The crystal structure of ecUGM was the first in this class of enzyme solved (Sanders et al, 2001). Subsequently, crystal structures of kpUGM and mtUGM were determined (Beis et al, 2005). UGM is a homodimer and belongs to the α/β class of protein. Each monomer of UGM has three distinct domains (Figure 1-7). Domain 1 is a FAD binding region that is made of αβα Rossmann fold, a unique secondary structural element/motif found in nucleotide binding proteins. Most of the interactions of FAD are

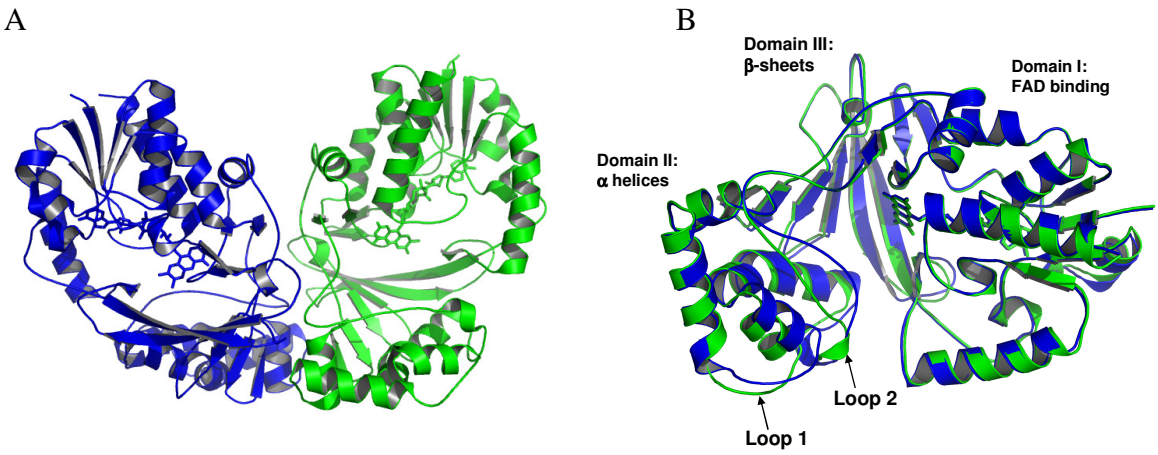


Figure 1-7 Structure of ecUGM. **A)** Ribbon diagram representation of the dimer, monomer A (blue) and monomer B (green). FAD represented as sticks. **B)** Overlay of monomer A and monomer B to highlight the differences in the mobile loop conformation. Note the closed and open conformation of the mobile loop in monomer A and monomer B respectively.

through hydrogen bond formation with conserved side chains. The two methyl groups of the isoalloxazine ring in FAD are stabilized by hydrophobic contacts with the conserved tyrosines (for example Tyr314 in kpUGM) The FAD binding domain (αβα Rossmann fold) is commonly found in other nucleotide binding proteins, but domain 2 and 3 were considered as unique for a FAD binding protein. Later, folds similar to domain 1 and 2 were identified in a FAD binding protein, known as protoporphyrinogen oxidase (PPO),

although, the function of UGM and PPO are unrelated (Corradi et al, 2006). Domain 2 is a α-helical bundle made of five α-helices with two mobile loops that can exist in an open and closed conformation (Sanders et al, 2001). Domain 3 is an anti-parallel β-sheet made of six β-strands. The majority of the active site residues are located in domain 2 and domain 3 (Chad et al, 2007; Sanders et al, 2001). The overall architecture of monomer A and monomer B is similar (rmsd of 1.4 Å for all equivalent C_α atoms), but differences in the mobile loop conformation were observed. In monomer A, the mobile loops exist in a closed conformation, while in monomer B an open conformation was observed for the mobile loop (Sanders et al, 2001). The overall structures of ecUGM, kpUGM and mtUGM similar and the monomers of ecUGM, kpUGM and mtUGM could be superposed with an rmsd of 1.0 Å (ecUGM and kpUGM), 1.2 Å (ecUGM and mtUGM), 1.2 Å (kpUGM and mtUGM) for all equivalent C_α atoms. The sequence identity between ecUGM and kpUGM, ecUGM and mtUGM, kpUGM and mtUGM are 39 %, 42 % and 37 % respectively.

1.5 Conformation of oxidized and reduced FAD

Crystal structures of ecUGM and mtUGM were determined with oxidized FAD. kpUGM structure was solved with both oxidized and reduced FAD (Beis et al, 2005). In oxidized UGM structures, the isoalloxazine ring of FAD is planar and N5 atom is sp^2 hybridized (Figure 1-5). In reduced structure (FADH-) the isoalloxazine ring is slightly puckered and N5 is sp^3 hybridized and it is known as a butterfly or bent conformation (Figures 1-5 and 1-8). In $FADH^-$ the puckering of isoalloxazine ring can occur in two ways, namely, *re*-face or *si*-face. In $FADH_2$ containing structures, like thioredoxin reductase (PDB code: 1C10), polyamine oxidase (PDB code: 1B37), and cholesterol

oxidase (PDB code: 1COY) *re*-face bending for isoalloxazine ring was observed (Binda et al, 1999; Haynes et al, 2002; Lennon et al, 1999; Yue et al, 1999). In reduced kpUGM structure, isoalloxazine ring of $FADH^-$ is puckered in a *si* conformation. The N5 atom in

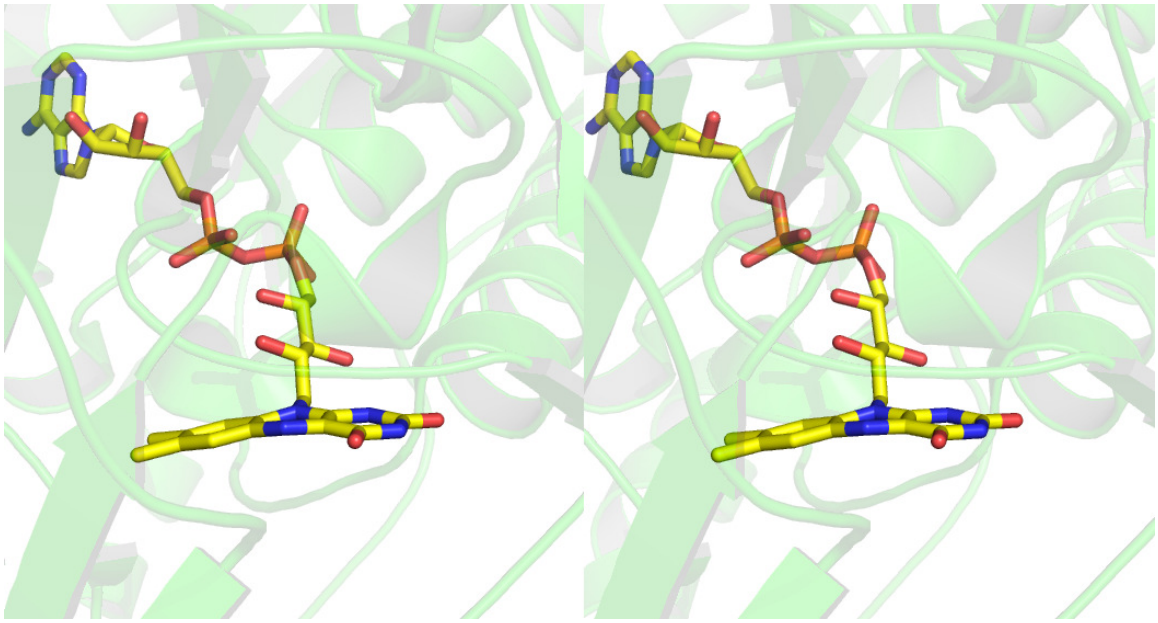


Figure 1-8 Structure of reduced kpUGM. Ribbon diagram representation (stereoview) of the reduced kpUGM (PDB code 2BI8). The N5 of $FADH^-$ (represented as sticks) is pointing towards the protein (*si* face).

the *si* conformation is not facing the substrate binding cleft, but point towards the protein (Figure 1-8). The *si*-face bending was also observed for the isoalloxazine ring of FMN in nitroreductase from *Enterobacter cloacae* (PDB code: 1KQD). The *si* conformation for FADH- in kpUGM causes significant difficulty for covalent catalysis to occur as this may cause steric clashes between the sugar moiety and active site residues. In order to avoid steric clashes and form an FAD-galactosyl adduct, the protein would need to undergo significant conformational changes including the backbone structure. It has been proposed that reduced kpUGM may favor the *si* conformation over the *re* conformation and substrate binding may perturb the active site architecture to favor *re* conformation

over *si* conformation. Also, the energy barrier for the interconversion between *re*-face and *si*-face bending is only 4-5 kcal/mol (Moonen et al, 1984), which could be gained upon substrate binding.

1.6 Role of active site residues

The crystal structures of ecUGM, kpUGM, mtUGM together with the sequence alignment of various bacterial UGMs, revealed the conserved residues involved in FAD and substrate binding (Chad et al, 2007). The role of active site residues in ecUGM and kpUGM have been studied through site-directed mutagenesis (Chad et al, 2007; Sanders et al, 2001). The active site mutants of kpUGM and their kinetic parameters are shown in Table 1-1. It is evident that most of the mutations affect the enzyme activity. Most importantly, two conserved arginines are critical for activity. A pronounced effect on substrate binding was also observed on replacing the active site tryptophan (W160A in kpUGM). Modeling studies suggest that Trp184 is involved in stacking interactions with the uracil ring of the substrate and this is consistent with the results of the mutagenesis experiments. The four active site tyrosines were found to be important for substrate binding, but, not critical for activity. The two acidic residues, Asp351 and Glu301 are conserved in prokaryotic UGM, while in some homologs the residue that corresponds to Asp351 is an asparagine and the residue that corresponds to Glu301 is an aspartic acid (Chad et al, 2007). Mutation of these acidic residues affects binding and catalytic activity. It is possible that these acidic residues may be involved in stabilization of a positively charged intermediate formed during the reaction or act as hydrogen bond donor/acceptor for the substrate. The two active site arginines (Arg174 and Arg280) are absolutely

essential for UGM activity, replacement of either of the arginine residue results in a completely inactive enzyme. Arg174 is located on a mobile loop away from the active

Table 1-1 Kinetic parameters for kpUGM active site mutants.

Protein	K_m (μM)	k_{cat} (min^{-1})	k_{cat}/K_m
WT	43 ± 6	330 ± 40	7.7×10^6
E301A	205 ± 18	18 ± 2	8.8×10^4
D351A	1002 ± 284	8 ± 2	8.0×10^3
W160A	*2.53×10^6	n/a*	130
R280A	ND	ND	
R174A	ND	ND	
Y185F	386 ± 92	78 ± 14	2.0×10^5
Y155F	619 ± 117	217 ± 25	3.5×10^5
Y349F	739 ± 63	103 ± 16	1.4×10^5
Y314F	819 ± 182	313 ± 40	3.8×10^5

*K_m value calculated assuming the k_{cat} would be similar to wild type
ND (not determined)

site and it is apparent that the flexibility of this loop could result in the movement of Arg174 towards the active site and stabilize the negatively charged phosphate groups of the substrate. Modeling studies support the movement of this flexible loop and the side chains of conserved arginine (Arg170 in ecUGM and Arg174 in kpUGM) make contacts with the phosphate groups (Chad et al, 2007; Yuan et al, 2008). Modeling studies also suggest that the β-phosphate group of UDP-Gal*p* can be stabilized by the side chain of Arg280 (in kpUGM). Stabilization of the phosphate groups of the substrate by the two conserved arginines underline their functional role in UGM.

1.7 Molecular simulations

In order to understand the binding mode of substrate and mobile loop movement modeling studies (docking and molecular dynamics simulation) have been carried out on UGM (Chad et al, 2007; Yuan et al, 2008; Yuan et al, 2005). AutoDock (Goodsell et al,

1996) was used to model the substrate binding into the active site. Docking studies were performed using ecUGM, monomer A and B. The monomers of ecUGM differ in their structure with respect to mobile loop conformation. The mobile loops exist in a closed and open conformation in monomer A and monomer B respectively. The mobile loops in kpUGM and mtUGM exist in a more open conformation. The substrate binding pocket of ecUGM monomer A and B are slightly different due to the open and closed conformation of the mobile loop. The docked conformation of UDP-Gal*p* in monomer A showed numerous interactions with the side chains of conserved active site residues. In this binding mode, the uracil ring of UDP-Gal*p* is involved in π-stacking interactions with Trp156 (ecUGM). The β-phosphate group of UDP-Gal*p* was stabilized by the side chain of Arg278. The sugar moiety is in close proximity to the isoalloxazine ring of FAD (3.0 Å from N5 of FAD to anomeric carbon of galactopyranose). The majority of interactions for UDP-Gal*p* are from the UDP moiety which is surrounded by 15 residues. This is consistent with the results obtained from saturation transfer difference (STD) NMR experiments that showed strong signals for the protons of UDP as compared to the galactose moiety (Yuan et al, 2005). STD-NMR experiments were done for UDP and its signals were similar to the UDP portion of UDP-Gal*p*. This suggests that uridine moiety play an important role in the binding of UDP-Gal*p* to UGM.

In monomer B, the docked conformation of UDP-Gal*p* is significantly different from monomer A. In this binding mode, the sugar moiety is not located adjacent to the isoalloxazine of FAD. Also, the stacking interactions between the uracil ring of UDP-Gal*p* and Trp156 are not observed. However, the salt-bridge interactions between the Arg278 and the phosphate group of UDP-Gal*p* in monomer B were similar to the one

observed in monomer A. Docking studies on UGM were helpful in predicting the binding mode of UDP-Gal*p* and also underline the role of some active site residues. To account for the flexibility of the enzyme, especially the movement of mobile loops, molecular dynamics simulation (MDS) were performed on UGM. MDS is one of the modeling techniques that can be used to simulate the movement of atoms in a molecule on a time scale and analyze the changes in the structure for the given time period. MDS calculations on ecUGM (monomer B) with modeled UDP-Gal*p* revealed the rearrangement of mobile loops, i.e., from open to closed conformation. Arg170 located in the mobile loop was found to interact with the phosphate group of UDP-Gal*p* (Yuan et al, 2008). Similar results from MDS studies were observed for kpUGM (Chad et al, 2007) and mtUGM (unpublished data). MDS studies corroborate the proposed mobile loop movement that occur upon substrate binding and is consistent with the mutagenesis studies of Arg174 of kpUGM (Chad et al, 2007).

1.8 Enzyme kinetics and inhibition

The work presented in this thesis includes enzyme kinetics and inhibition studies on UGM. Hence, an introduction to these two topics is given in this section. Enzymes are catalysts that enhance the rate (velocity) of a reaction without affecting the equilibrium of the reaction. In most cases, enzymes convert one chemical (substrate) into another (product). Some enzymes can bind to more than one substrate. Enzyme catalysts differ from chemical catalysts in many ways, for example, enzymes achieve remarkable rate enhancements, high degrees of specificity, diversity in the types of reaction catalyzed and tight regulation of catalysis by cellular metabolites (Kraut 1988). The active site of an enzyme is defined as the location on the enzyme where catalysis takes place. The rates of

enzyme-catalyzed reaction can be decreased or increased by inhibitors or activators respectively. Enzyme kinetics is an important experimental tool to study the rates of enzyme-catalyzed reactions. The overall goal of enzyme kinetics is to measure, analyze and interpret the rate of enzyme-catalyzed reactions. Enzyme kinetics provides evidence about the enzyme mechanisms and is also helpful in defining the role of active site residues involved in substrate binding and catalysis.

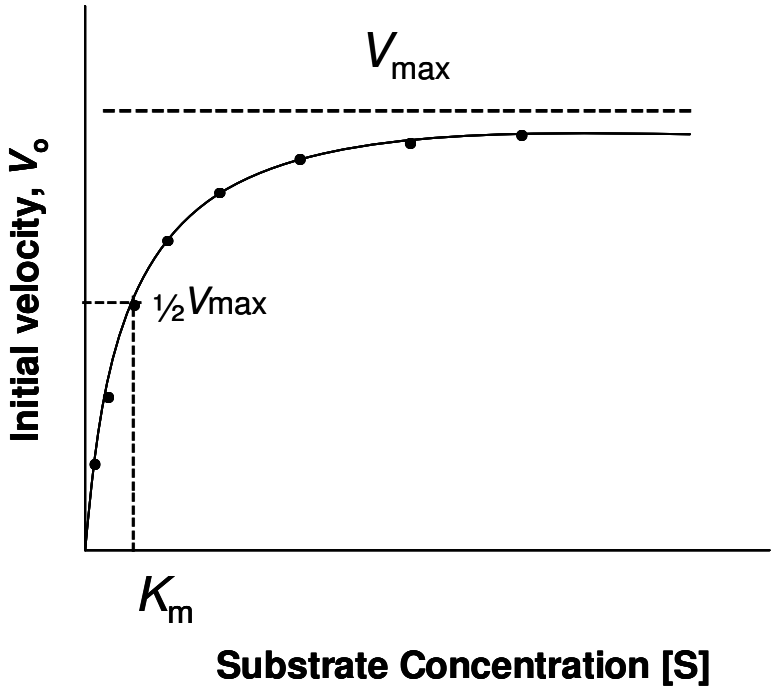


Figure 1-9 Effect of substrate concentration on the initial velocity of an enzyme-catalyzed reaction. The substrate concentration at which V_o is half-maximal is K_m, the Michaelis constant.

The rate of enzyme catalyzed reaction is influenced by many parameters and one key factor is the substrate concentration [S]. The effect on V_o (initial velocity) by varying [S] at constant enzyme concentration is shown in Figure 1-9. At relatively low [S], V_o increases linearly with increase in [S]. However, at higher [S], V_o increases slowly and reaches a point beyond which there is insignificant changes in the V_o with increasing [S].

This flat region (plateau) of the graph is called the maximum velocity V_{max}. The relationship between the substrate concentration and rate of enzyme catalyzed reaction can be expressed quantitatively using Michaelis-Menten equation (5-1).

$$V_o = \frac{V_{max}[S]}{K_m + [S]} \qquad (5\text{-}1)$$

K_m is the [S] at which V_o is one-half of V_{max}.

K_m can also be defined by the expression,

$$K_m = \frac{k_2 + k_{-1}}{k_1} \qquad (5\text{-}2)$$

k_1, k_{-1} and k_2 are shown in equation 5-3

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} E + P \qquad (5\text{-}3)$$

In equation 5-3, ES represents the enzyme-substrate complex and P represents the product. When $k_2 << k_{-1}$, then K_m reduces to k_{-1}/k_1, which is a dissociation constant (K_d) for the ES complex. Under these conditions K_m can be considered as a measure of the affinity of the enzyme for the substrate. For a reaction with multiple steps after formation of the ES complex, then K_m becomes a complex function of many rate constants. K_m and k_{cat} are the two widely used parameters in enzyme kinetics. k_{cat} is a first-order rate constant with units of reciprocal time and is also known as the turn-over number. It is the number of substrate molecules converted to product in a given unit of time by a single enzyme molecule when the enzyme is saturated with the substrate. In order to compare the catalytic efficiency of different enzymes or different substrate towards an enzyme the factor k_{cat}/K_m is generally used.

Inhibitors are molecules that bind to a specific site of an enzyme and prevent enzyme catalysis. The inhibitor may bind reversibly or irreversibly to the enzyme. The structure of the inhibitor may or may not resemble the substrate molecule, when inhibitor structure is similar to substrate then it is called as substrate mimic or substrate analog. Substrate analogs are often used as probes to investigate enzyme mechanisms (John, 2007).

Reversible inhibitors show three different types of inhibition mode, namely, competitive, uncompetitive and noncompetitive (Figure 1-10). Competitive inhibitors compete for the same site as the substrate and they bind only to free enzyme. The inhibitory activity of a competitive inhibitor can be altered by substrate concentration, for example, increasing the substrate concentration can displace the inhibitor from the active site and decrease the amount of inhibitor binding to the same site. If an inhibitor binds only to the ES complex, then it is termed as uncompetitive inhibition. In uncompetitive mode high concentration of substrate increase the concentration of ES and this will increase the amount of inhibitor binding to it. In case of noncompetitive inhibition, the inhibitor binds to either the free enzyme or the ES complex. Inhibitor binding does not affect the binding of substrate, i.e, substrate can still bind to the enzyme. The enzyme is inhibited by the inhibitor irrespective of the presence or absence of substrate. This mode of inhibition is also called as “mixed” inhibition and such inhibitors are termed as mixed inhibitors. The uncompetitive and noncompetitive modes of inhibition are often observed with enzymes that catalyze the reaction between two or more substrates. Most importantly, uncompetitive and noncompetitive inhibitors bind to a site different from the substrate binding site. The three modes of inhibition can be determined by various

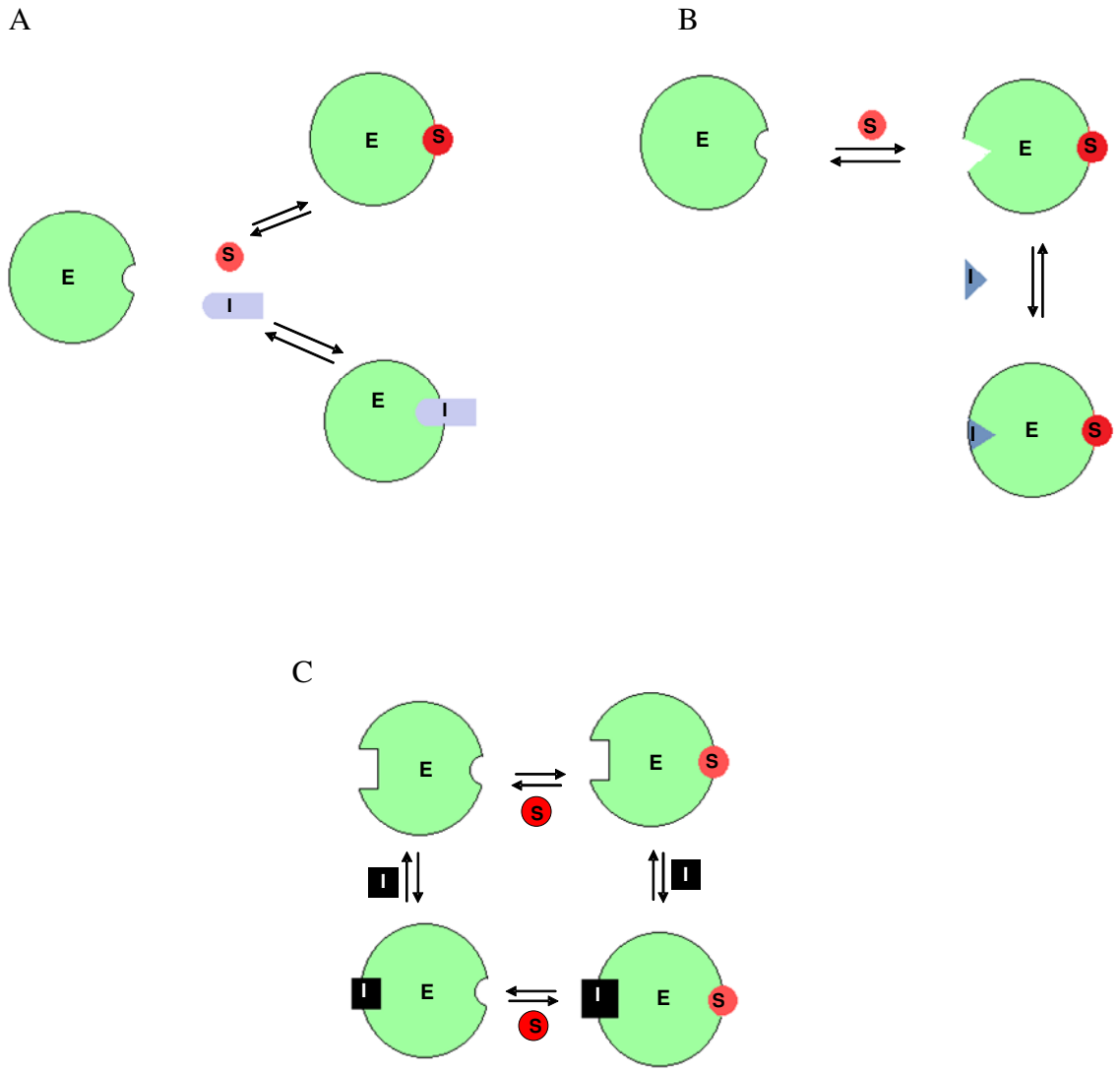


Figure 1-10 Schematic representation of the three types of reversible inhibition (E, enzyme; S, substrate; I, inhibitor). **A)** Competitive inhibitors bind to the active site of the enzyme **B)** Uncompetitive inhibitors bind at a distinct site to that of the substrate, but, bind only to the ES complex. **C)** Noncompetitive inhibitors bind at distinct site, but, can bind to E or the ES complex.

graphical plots (for example, Lineweaver-Burk plot) of the kinetic analysis carried out in the presence of inhibitor. The pattern of the lines from this plot can suggest the type of inhibition observed with the inhibitor.

1.9 Inhibitors of UGM

As discussed previously, UGM is a potential target for antibacterial drug development. Many research groups have designed, synthesized and evaluated diverse chemical scaffolds as inhibitors of UGM (Peltier et al, 2008; Richards & Lowary, 2009). Only a few of them showed promising inhibitory activity towards UGM. The inhibitors of UGM can be classified into distinct structural classes and for better clarity each class will be discussed separately.

1.9.1 Sugar-based derivatives

A number of sugar-derivatives have been synthesized as inhibitors of UGM (Figure 1-11). Among them are the imino-sugars (Desvergnes et al, 2007; Ghavami et al, 2004; Veerapen et al, 2004), alditolaminophosphonates (Pan et al, 2007) and pyrrolidine analogs (Lee et al, 1997) (Figure 1-11). The sugar-based derivatives showed no inhibition or poor inhibitory activity.

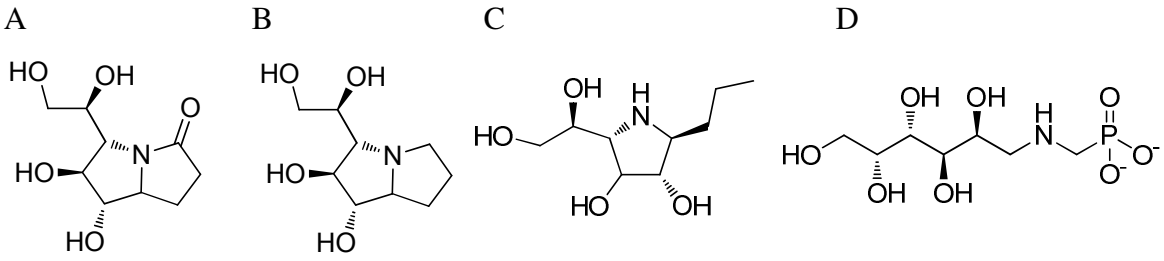


Figure 1-11 UGM inhibitors based on sugar derivatives. **A)** and **B)** iminosugars **C)** pyrrolidine analog D) alditolaminophosphonates.

The sugar derivatives may occupy only the sugar binding region of the active in UGM and do not make contacts with the residues of the UDP binding region. Modeling

suggests that the residues near the UDP region contribute to the binding of UDP-Gal*p* and absence of this moiety in these molecules may account for their poor inhibition.

1.9.2. Substrate analogs

Inhibitors based on substrate (UDP-Gal*p* and UDP-Gal*f*) have been synthesized as inhibitors of UGM (Figure 1-12). Notable among these substrate analogs are the fluorinated analogs of UDP-Gal*p* and UDP-Gal*f*. The 2-fluoro and 3-fluoro analogs of UDP-Gal*p* were found to be poor substrates of UGM, but neither of them were tested for their inhibitory activity against UGM (Barlow & Blanchard, 2000). The 4-fluoro analog of UDP-Gal*p* was shown to inhibit (~ 50 % at 166 μM) the conversion of UDP-Gal*f* to UDP-Gal*p* (Burton et al, 1997). 2- and 3-fluoro analogs of UDP-Gal*f* were synthesized and evaluated for activity against UGM. These analogs are also poor substrates for UGM, but are not inhibitors (Zhang & Liu 2001).

Sugar nucleotide mimetics such as iminogalactofuranose-UMP conjugates were synthesized as inhibitors of UGM, but the biological activity of these compounds were not reported (Liautard et al, 2006). A peptide analog of UDP-Gal*f* based on iminogalactofuranose template has been designed and synthesized as inhibitor of UGM (Lee et al, 1999), although the inhibitory activity for this compound is not known. Phosphonate analogs of UDP-Gal*p* and UDP-Gal*f* were synthesized as inhibitors for UGM (Caravano & Vincent, 2009). These analogs were evaluated for activity (towards ecUGM) under non-reducing and reducing conditions (See section 5.3.1). The phosphonate analog of UDP-Gal*p* showed 36 % inhibition at 1 mM under non-reducing condition, but under reducing conditions < 10 % inhibition was observed. UGM is active

only when it is reduced and it is more appropriate to perform inhibition studies under reducing condition to evaluate their potency. The substrate (UDP-Gal*f*) concentration in

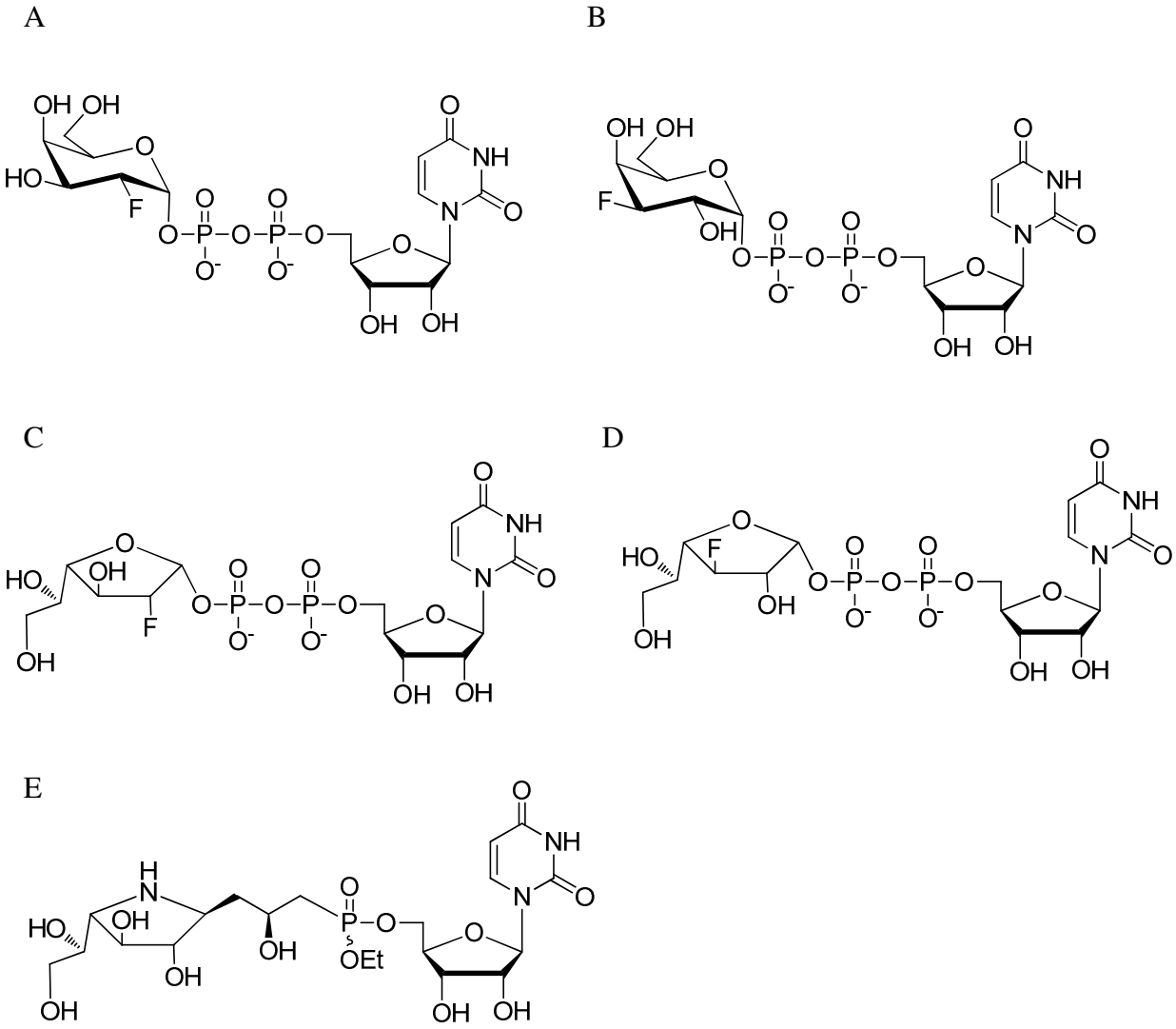


Figure 1-12 Substrate analogs as UGM inhibitors. **A)** and **B)** Fluoro analogs of UDP-Gal*p* **C)** and **D)** Fluoro analogs of UDP-Gal*f* **E)** iminogalactofuranose-UMP conjugate.

their assay was set to 1 mM, while the K_m for ecUGM is only 27 μM (Caravano & Vincent, 2009). The higher concentration of substrate used in the assay may affect the proper evaluation of inhibitors (See section 5.3.1).

1.9.3. Thiazolidinone derivatives

Inhibitors based on 5-arylidene-2-thioxo-4-thiazolidinone (ATT) core were identified in a high throughput screening (HTS) against kpUGM (Carlson et al, 2006). Based on the ATT core various analogs have been synthesized and evaluated for inhibitory activity against kpUGM and mtUGM. One compound (Figure 1-13) in this series showed promising activity towards kpUGM (IC_{50} = 1.6 μM), but, its activity towards mtUGM is 40-fold lower (IC_{50} 65 μM). These compounds showed selectivity

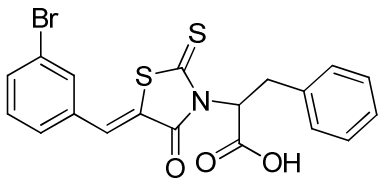


Figure 1-13 UGM inhibitor based on 5-arylidene-2-thioxo-4-thiazolidinone scaffold.

towards kpUGM as compared to mtUGM. It is possible that subtle differences in the active site architecture between mtUGM and kpUGM may account for the differences in activity. Unfortunately, these compounds have the possibility of undergoing covalent modification (due to the conjugated system) by reacting with nucleophiles like thiols and this is an undesirable property in a lead like molecule for further development.

1.9.4. Nitrofuranyl amides

Antibacterial agents based on a nitrofuran scaffold are well known in the pharmaceutical market, for example, nitrofurantoin used in the treatment of urinary tract infections. Also, the antimicrobial agent metronidazole has structural similarity to a nitrofuran scaffold. Based on this precedent, a library of compounds based on nitrofuranyl amides were designed, synthesized and evaluated for their inhibitory activity

towards mtUGM (Figure 1-14) (Tangallapally et al, 2004). Compounds from this series showed moderate activity in *in-vitro* assays. The MIC activities of these compounds are comparable to the MIC values of front-line anti-tuberculosis agents, like, isoniazid and ethambutol. The moderate IC_{50} values of these compounds do not correlate with the excellent MIC values and this suggests that targets other than UGM may account for their potent MIC activities.

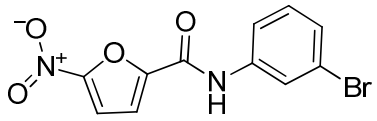


Figure 1-14 UGM inhibitor based on nitrofuran core.

1.9.5. Uridine derivatives

A microtiter plate-based assay has been designed to identify inhibitors of UGM (Figure 1-15) (Scherman et al, 2003). The assay is based on the formation of tritiated formaldehyde from UDP-Gal*f* and not from UDP-Gal*p* in the presence of periodate. This assay was successfully used to screen a uridine-based chemical library and a potential hit molecule was identified (320KAW73). The approximate IC_{50} value of 320KAW73 was found to be 6.0 μM, but the inhibitor was found to be inactive in a cell-based assay.

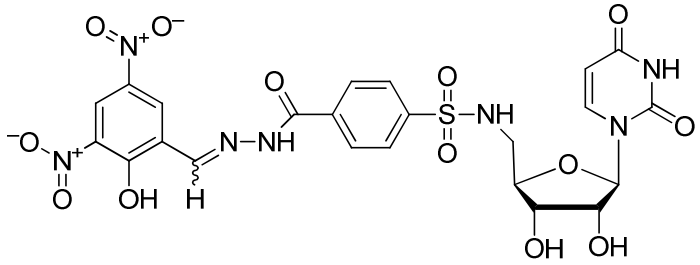


Figure 1-15 Uridine based inhibitor identified through microtiter plate-based assay.

1.10 Research Objectives

The unliganded UGM structures (ecUGM, kpUGM and mtUGM) provided insights onto the UGM structure, FAD binding and also the role of few active site residues. The crystal structure of UGM bound to substrate was not available or reported at the time of initiating this study. One of the objectives of the current study is to determine the structure of UGM in complex with the substrate (UDP-Gal*p*) by X-ray crystallography.

UGM:substrate complex structures

The UGM:substrate complex structure would help us to understand the active site interactions and the role of each amino acid at a molecular level. UGM:substrate complex structures would reveal the conformational changes (mobile loop rearrangements) that occur upon substrate binding and its effect on interactions with the substrate. Also, UGM:substrate complex structures may provide a platform for structure-assisted design of inhibitors for UGM. The structures of oxidized and reduced UGM:substrate complexes would reveal the structural differences between them and the role of FAD. In addition, I also address the question of how UMP, UDP, and UDP-glucose bind to UGM.

Inhibition studies on UGM

Inhibitors of UGM based on various chemical scaffolds have been reported, but most of them have only moderate inhibitory activity. Another objective of this thesis is to identify UGM inhibitors that could act as antibacterial agents. Both ligand and structure-based approaches were used to identify inhibitors for UGM.

CHAPTER 2

Outline of Protein crystallography

2.0 Introduction to protein crystallography

The biological function of a protein molecule and its three dimensional structure are closely related, i.e., structure determines the function. Protein structures can be determined by NMR or X-ray crystallography (Wlodawer et al, 2008). Scientific and technological advances in NMR and X-ray crystallography have led to the rapid determination of three dimensional structures of various macromolecules. The structures reported here are determined by X-ray crystallography and this chapter is an overview of the various steps involved in protein crystallography. Protein crystallography essentially involves four major steps that include: 1) Crystallization, 2) Crystal diffraction (Data collection and processing), 3) Phase determination and 4) Refinement and model-building.

2.1 Crystallization of proteins

Crystal growth is a phase transition phenomena to form solid particles by aggregation of molecules in the solution (McPherson, 2004; McPherson, 2009). The process of aggregation does not occur in unsaturated or saturated solutions. In the case of undersaturated systems, the concentration of the solute is below its solubility limit, however, when the concentration of the solute reaches its solubility limit the system is saturated. In a saturated system, equilibrium is established between the solid and the solution state, but crystals cannot grow under these conditions. A supersaturated solution represents a system where the concentration of the solute exceeds its solubility limit, i.e,

more solute molecules are present in solution than its solubility limit and such a system is not in equilibrium. Hence, a supersaturated system is essential for crystal growth.

A plot of precipitant concentration vs protein concentration (phase diagram) reveals different zones that are favorable and unfavorable for crystal growth (Figure 2-1).

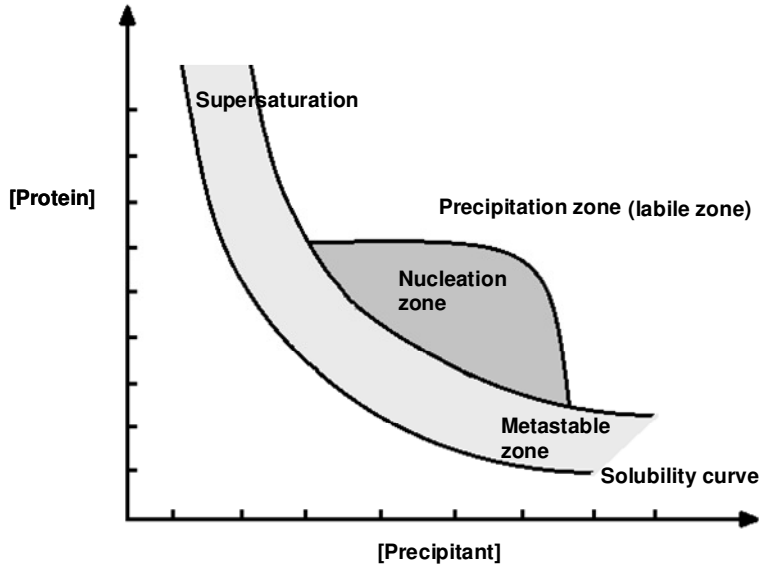


Figure 2-1 Schematic illustration of protein crystallization phase diagram.

The process of crystal growth can be divided into three stages, namely, nucleation, crystal growth and cessation of growth. Unsaturated and saturated systems are unfavorable zones in the phase diagram that do not result in crystal growth. When the system is supersaturated then spontaneous formation of critical nuclei (labile zone) occurs and this triggers the aggregation of molecules leading to the formation of crystals. The metastable zone in the phase diagram represents another supersaturated system ideal for crystal growth, although nuclei formation does not occur in this region. The labile region (precipitation zone) not only forms crystal nuclei, but amorphous precipitate as well and this is a major problem in protein crystallization. In order to form protein

crystals, one has to bring the system to the labile zone to form few nuclei which will continue to grow in the metastable zone without formation of any new nuclei and precipitation. The termination or cessation of crystal growth could be due to drop in supersaturation levels, crystal surface defects and impurity poisoning that inhibits the crystal growth.

Proteins can be crystallized by different methods (Table 2-1) and all of them have a common theme of bringing the protein solution to a supersaturated state, which leads to nucleation and crystal growth. Among the various methods for crystallization, vapor diffusion methods are more popular than any other approach.

Table 2-1 Various methods for protein crystallization.

S. No	Method
1	Bulk crystallization
2	Batch method
3	Dialysis
4	Seeding
5	Free interface diffusion
6	Vapor diffusion
7	Temperature induced

Vapor diffusion methods are of two types, sitting drop and hanging drop. If the drop (protein and precipitant solution) is supported by some surface, it is called sitting drop, and if the drop is suspended from some surface it is called hanging drop (Figure 1-10). The principle in both methods is essentially the same. The protein/precipitant solution is allowed to equilibrate in a closed system and this will increase the precipitant and protein concentration (in the drop) driving the system to supersaturation and crystal formation.

Microbatch crystallization is a variant of batch crystallization carried out under oil (Chayen, 1997; Chayen, 1998). The protein solution and crystallization reagent are pipetted under a small layer of oil (Figure 2-2). The oil layer acts as a barrier and hence there is little or no diffusion of water through the oil. The major advantage of batch or microbatch method is all the reagents involved in the crystallization are present at a specific concentration to create a supersaturated system. Also, in batch experiments

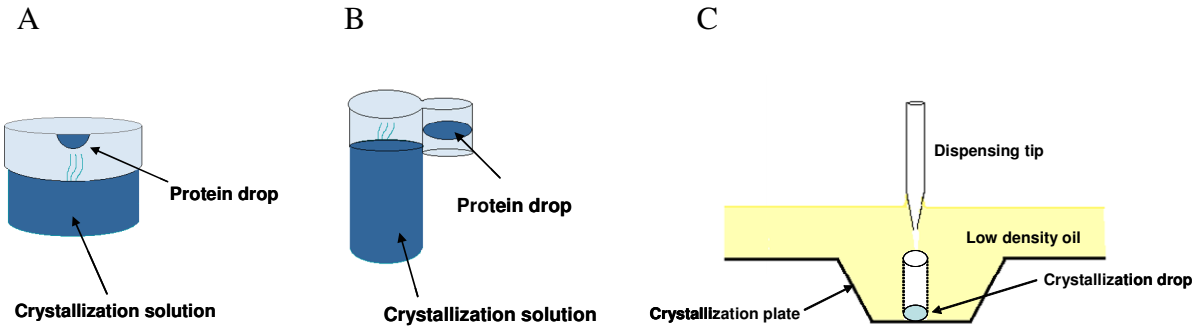


Figure 2-2 Schematic representation of the set up for vapor diffusion and microbatch crystallization. **A)** Hanging drop **B)** Sitting drop **C)** Microbatch method.

the concentration of the protein and the crystallization reagent in the drop do not vary significantly over the period of time. Microbatch methods are well suited for automation and require less volume of protein and crystallization reagents.

2.2 Crystallization of protein-ligand complexes

Crystals of protein-ligand complexes can be formed by two different approaches, namely, cocrystallization and soaking (Hassell et al, 2007). A major difference between them is the time at which ligand is incorporated into the protein. In cocrystallization, the ligand is added to the protein solution to form protein-ligand complex and then crystal trials are carried out using this protein-ligand solution. Soaking methods rely on preformed protein crystals and the ligand is introduced by soaking the crystal into a

crystallization reagent that contains the ligand at appropriate concentration. A number of parameters affect the incorporation of ligand by cocrystallization and soaking as well. For example, cocrystallization is sensitive to ligand concentration and solubility, protein concentration and temperature. The ligand solubility is a major factor in forming protein-ligand complexes. To increase the chances of forming protein-ligand complexes, higher concentration of ligand is required without any precipitation when mixed with the protein solution. In general, the ligand concentration that is required to form protein-ligand complex is 10 times more than the dissociation constant (K_d) of the ligand to achieve full occupancy. The solubility of ligand may vary with respect to temperature; thus protein-ligand complex formation by cocrystallization can be carried out at various temperatures. The addition of a ligand to a concentrated protein solution may result in the precipitation of protein due to the limited solubility of the ligand. One way to overcome this issue is to mix the diluted protein sample and dilute ligand solution, for example crystals of kinase 4 were formed by diluting the protein to 1 mg/ml and diluted ligand added (at 1:3) to form a stable complex and found to give well diffracting crystals.

Soaking is a common method of choice to obtain protein-ligand complex crystals (Hassell et al, 2007). A number of factors need to be taken into consideration in forming protein-ligand crystals by soaking. Most important parameter is the stability of protein crystals. Crystals should be stable and withstand the soaking conditions. Conformational changes that occur upon soaking may result in cracking of crystals and eventually dissolve them. The conformational changes observed with soaking may not represent the true conformational state of the protein and could be validated with crystals obtained by cocrystallization experiments. The soaking time and the concentration of the ligand needs

to be optimized to achieve full occupancy. It has been shown that addition of certain compounds, known as additives, stabilize the crystals during the soaking process (Hassell et al, 2007). Use of additives is also common in the cocrystallization method. Ligands can also be soaked into preformed protein-ligand complex crystals and this process is known as ligand exchange. The success of this method depends on the affinity (K_d) of the second ligand to successfully displace the preoccupied ligand.

2.3 Crystal harvesting and mounting

Crystals from the crystallization drop need to be isolated (harvested) and mounted on a goniometer for diffraction experiments (Pflugrath, 2004). Crystals are looped from the drop and placed in the crystallization reagent that contains a chemical or compound known as cryoprotectant. When protein crystals are exposed to X-rays, free radicals are generated that can damage the protein crystals. The damage by free radicals can be minimized by collecting the data at cryogenic temperatures (-196 ^{o}C) by using a stream of nitrogen gas. This is known as cryomounting. Under these conditions ice crystals are formed that are detrimental to protein crystals. The presence of a cryoprotectant in the crystallization reagent favors the formation of glassy ice over ice crystals. A wide range of compounds are available for cryoprotection, such as glycerol, ethylene glycol, sugars, PEGs etc. Crystals can be mounted using a small loop made from a thin fiber, but different types of loops are are available to meet the specific requirements (Hampton Research Inc).

2.4 Why do we use X-rays?

If one wishes to determine an image of a protein molecule at an atomic level, then a light source (electromagnetic wave) of suitable wavelength must be used. The atomic

structure of a molecule can be obtained if the wavelength of the electromagnetic radiation is 1-2 Å (interatomic distance in protein molecules). In the electromagnetic spectrum, X-rays match this criteria and their wavelengths range from 2.3 - 0.6 Å (5-25 KeV). Hence, X-rays are used to study the three dimensional structures of protein molecules that are packed in a crystal (Rhodes 2006; Rupp 2010).

X-rays that are useful for crystallography can be produced by bombarding a metal target (Molybdenum or Copper) with electrons. When high-energy electrons strike the metal atom it displaces an electron from the lower orbital and then an electron from a higher orbital drop into the vacant orbital, releasing X-ray photons of particular energies, known as characteristic radiation. There are three common X-ray sources, namely, X-ray tubes, rotating anode tubes and particle storage rings (synchrotron radiation). In case of X-ray tubes heat dissipation from the anode by the circulating water affects the output. A higher X-ray output can be obtained by using rotating anode tubes, in which the target metals rotate rapidly and this arrangement helps in heat dissipation by spreading the bombardment over the large piece of metal. In particle storage rings, charged bodies like electrons emit energy (synchrotron radiation) when forced into a curved motion and the energy emitted include X-rays. They are the most powerful X-ray sources.

2.5 Crystals, diffraction and symmetry

Cystals are regular three dimensional arrays of atoms, groups, molecules or molecular assemblies (Rhodes 2006; Rupp 2010). W. L. Bragg proposed that diffraction could be considered as a reflection from a set of equivalent parallel planes of atoms (electrons) in the crystals. Constructive interference (measurable reflections) from the

scattered rays will take place if the path difference between these rays is equivalent to an integral number of wavelength and this is known as Bragg's law (equation 2-1).

$$n\lambda = 2d \sin \theta \qquad (2\text{-}1)$$

n is an integer, λ = wavelength of the X-rays, d = interplanar spacing between the planes and θ = angle between the incident ray and scattering planes

The reflections or diffraction pattern of a lattice is also a lattice and its dimensions have reciprocal relationship to the real lattice. This arrangement of lattice points in reciprocal space is called the reciprocal lattice. The sphere of reflection or Ewald's sphere is an imaginary geometric construction that tells us how the reciprocal lattice point must be arranged with respect to the X-ray beam in order to satisfy Bragg's law and produce a reflection from the crystal. Considering a sphere with radius $1/\lambda$ (wavelength in reciprocal space) from the center of the crystal and any reciprocal lattice point in contact with the surface of the sphere will satisfy Bragg's law in reciprocal space and reflection occurs. Hence, during data collection it is necessary to rotate the crystal so that as many as possible reciprocal points come in contact with the surface of the sphere of reflection resulting in measurable reflections.

Individual atoms from a protein molecule can diffract X-rays, albeit too weakly to measure. A crystal can diffract X-rays strongly as compared to a molecule, because a crystal is composed of many identical units called unit cells. Each unit cell is composed of the same number of molecules packed in identical orientation. This arrangement of molecules in each unit cell can diffract X-rays in unison and results in the amplification of diffraction by each atom in a measurable form known as spots or reflections. The diffraction pattern observed for a protein crystal is characteristic and depends on the crystal symmetry. In order to determine the structure of a protein molecule, it is important

to know the symmetry information of the crystal, for example, the unit cell symmetry guide us to collect only the unique reflections and greatly reduce the magnitude of data collection. Based on the lattice symmetry it is possible to have seven crystal systems. Bravais showed that there are only 14 possible ways that identical lattice points can be arranged in space with the condition that each point has same number of neighbors at the same distances and in the same directions. These are called the 14 Bravais lattices (Rhodes 2006).

A symmetry operation is an event of converting one object into another identical object. Symmetry operations can take place about a point, a line or a plane. The point, line or plane about which symmetry operations are done is defined as a symmetry element. The combination non-translational symmetry elements passing through a single point (geometric centre of the crystal) lead to 32 crystal classes or point groups and crystals belong to any one these crystal classes. When symmetry operations are applied to points arranged periodically on a crystal lattice, the result is a space group. Two additional symmetry elements with translation are possible, namely, screw axes and glide planes. The combination of the 14 Bravais lattices with the 32 crystal classes and the possibility of translational symmetry (screw axes and glide planes) result in 230 unique crystallographic space groups. Biological macromolecules such as proteins (composed of only L-amino acids) crystallize in space groups that do not possess either inversion symmetry or symmetry plane (symmetry elements that change the hand of the molecule) and hence only 65 distinct space groups are possible for proteins. In addition to the unit cell symmetry, the diffraction pattern has a symmetry element called centre of symmetry or point of inversion at the origin. Friedel's law states that the intensity of reflections that

comes from opposite sides of the same set of planes are identical (equivalence of I_{hkl} and $I_{h\text{-}k\text{-}l}$). This additional symmetry in the reciprocal lattice allows further reduction in the total angle of data collection. An examination of diffraction pattern of a crystal may reveal that some reflections are missing (intensity of zero) and these missing reflections are due to certain symmetry elements present in the crystal. The reflections that are missing along the specific axis or planes are known as systematic absences. Systematic absences can assist in the determination of possible space group by looking at few sets of reflections.

2.6 Data collection

The data collection process involves the exposure of a crystal to X-rays and recording the diffraction pattern (the images) on a detector (Dauter, 1999). It is necessary to collect a few images to determine the crystal symmetry, unit cell dimensions, crystal orientation and resolution limit. This prescreening is essential to devise the strategy for data collection, which will maximize the completeness and the resolution of the dataset. The most common method of data collection in protein crystallography is the rotation or oscillation method. In the oscillation method, the crystal is rotated around a single axis in small increments, (0.1 to 1.5^{o}) and diffraction images are recorded for each of these small angles of rotation. The angle of rotation required to collect a complete data set depends on the crystal symmetry. The first few snapshots will give an indication on the extent to which the crystals diffract (resolution), the optimal exposure time, incremental angle, data anisotropy, the presence of ice rings, mosaicity, spot separation and sharpness of spots.

2.7 Data processing

Data processing involves three distinct steps:

1. Determination of the crystal cell parameters, space group and orientation.

2. Integration of the images with simultaneous refinement of crystal, beam and detector parameters (Leslie, 2006).

3. Data reduction to place all data on a common scale. Multiple observations are merged to give a unique dataset and outliers rejected. Intensities are reduced to structure factor amplitudes.

The step 1 involves the determination of the crystal unit cell dimensions and identifying the space group symmetry of the crystal (Kabsch, 2010a; Rossmann & van Beek, 1999). This process is known as indexing. Most of the indexing methods are based on the Fourier analysis of a subset of spot positions of one or more images. In step 2, the position of reflection in each image is predicted and the intensity value with estimated error for each observation of reflection is carried out by integration of the diffraction spot (signal) and subtracting it from the X-ray background (noise). In step 3, the intensities of reflections on each image are put on a common scale (scaling). This is necessary to account for the variations of the source intensity (e.g. beam decay at a synchrotron source), radiation damage and some absorption effects. Polarization and Lorentz correction factors are applied to account for these variations to put the data on a common scale. Once the data is scaled, multiple observations of the same reflections can be merged/reduced to a weighted mean intensity and standard deviation. At this stage of data processing various statistics are calculated to assess the data quality, such as R_{sym},

completeness of data, signal-noise ratio [I/σ(I)] etc. R_{sym} shows the agreement between the symmetry related reflections between various images. The lower the R_{sym} value (overall and high resolution shell) the better the data quality. The final step in the data processing is to convert the intensity data to structure-factor amplitude. The structure-factor has all the information (e.g., intensity of reflection) except the phase angle, the most critical data required to solve the protein structure.

2.8 Protein structure determination: The Phase problem

The aim of a protein crystallography experiment is to obtain the image of unit cell contents in the form of an electron density map (Rhodes 2006; Rupp 2010). The electron density map can be computed from intensities (from structure factors) and the phase angles of reflections using equation 2-2.

$$\rho(x,y,z) = \frac{1}{V} \sum_{h}\sum_{k}\sum_{l} |F_{hkl}| \, e^{-2\pi i(hx+ky+lz-\alpha'_{hkl})} \qquad (2\text{-}2)$$

In equation 2-2, ρ(x,y,z) is the electron density, *F(h, k, l)* represent the structure factor amplitudes and α_{hkl}, the phases. In X-ray crystallographic experiments, we can measure the intensities and the wavelength of diffracted rays, but, the phase angle information of the diffracted beam cannot be measured and this is known as the phase problem in protein crystallography. The phase information required to compute the electron density can be derived by various methods. The phase problem can be solved by using a protein structure that is similar to the unknown structure (sequence identity ≥ 30 %) and this is known as molecular replacement method (MR). Alternatively, the phases can be derived experimentally by incorporating heavy-atoms into the protein molecule or protein crystal and changes in the intensity pattern due to the heavy atom incorporation can be used to

solve the phase problem. The structures reported here were determined by the MR method and it will be discussed in greater detail.

2.9 Molecular replacement (MR)

Nearly three-quarters of structures deposited in the protein data bank (PDB) are solved by the MR method (Evans & McCoy, 2008; Rossmann, 2001). MR is helpful if the unknown protein structure is similar to a known structure and has ≥ 30 % sequence identity (Barton, 2008). The underlying assumption in MR is that the unknown protein structure will be similar to that of the known protein structure. MR avoids the time consuming process of heavy atom derivative preparation for solving the phase problem by isomorphous replacement or anomalous dispersion methods. Also, MR will be the method of choice for solving protein-ligand complexes if the structure of the unliganded protein is known.

In MR, the known protein structure that is used to calculate the initial phase of the unknown structure is called the phasing model. The phasing model has to be placed into the unit cell of the unknown protein and then used to estimate the initial phases for the properly positioned molecule. Placing the phasing model in the unit cell of the unknown protein in the right orientation and location is an important task in the MR method. The search procedure involves six variables, 3 rotation angles (Eulerian angles) for orientation and 3 translational parameters to find the position. In theory, it is possible to conduct all combinations of orientations and locations and compare the calculated amplitudes with the observed amplitudes obtained from the intensity data of the unknown protein. This complexity can be reduced by separating the search process for best orientation from the search for the translation or best position. A Patterson map, a vector map, is used to find

the best orientation and position of the search model within the unit cell of unknown protein. To find the best orientation, the rotation function uses intramolecular vector maps (self-Patterson peaks), as they depend only on the orientation and not on its position in the unit cell. If the model is properly oriented within the unit cell of the unknown protein then the Patterson map for the model and the unknown protein will be very similar. If the best orientation is found, then the particular solution can be used in the search for the best location/position of the model (translation). To find the best position, the translation function uses intermolecular vector maps (cross-Patterson vectors) as they depend on the position of the unit cell. A comparison of the Patterson maps between the model and that of the unknown protein would indicate the best position. If a best match is observed for a particular orientation and position, then the computed phases serve as the starting phases for the structure determination of the unknown protein. The search process is monitored by using R-factor (Equation 2-3), a parameter that compares the expected structure-factor amplitudes from the solution and the actual structure-factor amplitudes obtained from the intensity data of the unknown protein.

$$R = \frac{\Sigma \left| \left| F_{obs} \right| - \left| F_{calc} \right| \right|}{\Sigma \left| F_{obs} \right|} \qquad (2\text{-}3)$$

In general, R-values of 0.3-0.4 considered adequate as initial estimates of phases.

2.10 Refinement, model building, validation and deposition

Refinement is an iterative process of improving the model to agree with the experimental data, and the process of improvement is indicated by various statistics (Tronrud, 2004). The atomic positions, occupancies and temperature factors are the

parameters taken into consideration for comparing the calculated structure factor amplitudes and observed structure factor amplitudes. During the process of refinement additional constraints (fixed value for a parameter) and restraints (values specified within the range) are applied to achieve a near global minimum (better agreement between the calculated and experimental structure factor amplitudes). In addition, molecular mechanics based refinement such as molecular dynamics and simulated annealing can be used to find the lowest energy conformation and this should agree with the observed diffraction data. During the iterative process of refinement, the observed and calculated structure-factor amplitudes converge and improve the quality of the model. The most widely used measure of convergence is the residual index or R-factor. An initial model with an R-factor of 0.4 is promising and is likely to improve with the refinement process. Another statistical parameter that is used to indicate the quality and improvement of model is the free R-factor or R_{free}, and this is computed using small set of randomly chosen reflections (~ 5 %) that are not used during the refinement process. This is a cross-validation method or quality control process to assess the agreement between the calculated (from the model) and observed data. R_{free} measures how the current model predicts a subset of measured reflections not included in the refinement, and R-factor measures how well the current model predicts the entire data set that produced the model. In addition, proper care must be taken that the model obeys the chemical rules (expected bond angle and bond lengths), stereochemistry (no inverted centers of chirality) and conformational criteria, such as, planarity of peptide bond, trans- conformation of non-proline peptide bonds (except cis-proline), backbone conformational angles, Φ (Phi, N-C_α) and Ψ (Psi, C_α-C) are within the allowed regions by interpreting Ramachandran plots

(sterically allowed conformations for certain Φ, Ψ values) and the side chains exist in stable staggered conformations. The final model is validated through the PDB (Protein Data Bank) or Molprobity servers before deposition (Berman et al 2000; Davis et al 2004). After validation the model can be deposited in the PDB and each entry in the PDB is assigned a unique four letter code known as PDB accession code.

CHAPTER 3

Materials and Methods

3.0 Primers for PCR

1. drUGM N372D mutant

Forward 5' - CTA CCG CTA CTA CGA CAT GGA CCA GGT GG - 3'
Reverse 3' - CCA CCT GGT CCA TGT CGT AGT AGC GGT AG - 5'

2. kpUGM D351N

Forward 5' - CCT ACC GTT ACC TTA ATA TGG ATG TGA CCA TCG CC - 3'
Reverse 3' - GGC GAT GGT CAC ATC CAT ATT AAG GTA ACG GTA GG - 5'

3.1 Site-directed mutagenesis by PCR amplification

The drUGM N372D and kpUGM D351N mutants were prepared by PCR based site-directed mutagenesis procedure using Stratagene (now Agilent) QuickChange tool kit. The plasmid DNA from the wild type was used as the template. The primers for site-directed mutagenesis were designed using the program, MutaPrime^{TM}. Primers were obtained from AlphaDNA (AlphaDNA Inc). The components for PCR amplification and the protocol are shown in Table 3-1 and 3-2. Reactions were performed using a PTC-100^{TM} Programmable Thermal Controller (MJ Research Inc). After the reaction, the PCR product was digested with *DpnI* and transformed into *E. coli* NovaBlue cells. The cells were then plated and incubated at 37 ^{o}C overnight. Colonies were selected and inoculated into 10 ml of Luria-Bertani (LB) media and grown overnight at 37 ^{o}C. The cells were harvested and the cell pellet was used to perform plasmid DNA isolation using Qiagen plasmid isolation kit (QiaPrep). The concentration of isolated plasmid DNA was determined by a NanoDrop ND-1000 spectrophotometer (Thermo Scientific Inc) and the presence of the desired mutation was confirmed by DNA sequencing (Plant

Biotechnology Institute, Saskatoon). Sequencing results were analyzed with the program DNAMAN (Lynnon Biosoft Inc).

Table 3-1 PCR reaction components for site-directed mutagenesis.

S. No	Reaction components	Quantity
1	Template DNA	50 ng
2	Forward primer	125 ng
3	Reverse primer	125 ng
4	dNTP	1 µl
5	Reaction buffer	5 µl
6	*Pfu Turbo* DNA Polymerase	1 µl
7	RNAse free water	50 µl (Final volume)

Table 3-2 Reaction conditions for PCR.

Cycles	Segment	Temperature	Time
1	1	95 ^{o}C (Denaturation)	30 seconds
16	2	95 ^{o}C (Denaturation)	30 seconds
	3	55 ^{o}C (annealing)	1 minute
	4	72 ^{o}C (extension)	14 minutes (2 min/kb)
1	5	25 ^{o}C (cooling)	2 minutes

3.2 Determination of protein concentration

Protein concentrations were determined by colorimetric method using the Bradford assay (Bradford, 1976). The cationic and anionic forms of the CoomassieR

Brilliant Blue G-250 dye has different absorption maximum (λ_{max}) depending on the acidity of the solution. Under acidic conditions the anionic form (red) of the dye binds to the positively charged regions of the protein and the solution turns blue (cationic form of the dye). The absorbance of the resulting solution measured at 595 nm. Bovine serum albumin (BSA) was used as the standard. A series of dilutions were made from a BSA stock solution (10 mg/ml) to construct a standard curve. The concentration of unknown sample was determined from the regression equation obtained from the standard curve. All measurements were carried out using a Varian Cary 50 BIO UV-Visible spectrophotometer at 595 nm.

3.3 Expression and purification of drUGM WT

Previously, the gene encoding drUGM has been cloned (in our laboratory) into a pEHISTEV vector with an affinity tag at the N-terminus (six histidine residues). The construct (pEHISTEV vector with the drUGM gene) was transformed into *E. coli* Tuner cells (Novagen, USA). Transformed cells were grown in LB media with 50 μg/ml of kanamycin at 37 ^{o}C, until the optical density reached ~ 0.6 (at 600 nm), followed by induction with 0.4 mM isopropyl-β-thiogalactoside (IPTG) at 30 ^{o}C for 4-5 hours. Cells were harvested by centrifugation for 20 minutes (8000 rpm at 4 ^{o}C) and the resulting cell pellet was stored at -80 ^{o}C until purification. The frozen cell pellet was resuspended in 35 ml of lysis buffer (100 mM potassium phosphate pH 8.0, 1 mM AEBSF (4-(2-aminoethyl)-benzenesulfonyl fluoride), Triton-X 0.1 % v/v, DNase and lysozyme 20 μg/ml) and stirred at 4 ^{o}C for 30 minutes, followed by sonication. The suspension was clarified by centrifugation at 18000 rpm for 30 minutes. The supernatant was subjected to heat denaturation at 55 ^{o}C for 10 minutes, followed by centrifugation at 18000 rpm for 30

minutes. The supernatant was dialyzed against 25 mM potassium phosphate pH 8.0 (4 changes). The dialyzed sample was filtered and applied to a HQ20 (Applied Biosystems, USA) anion-exchange column, pre-equilibrated with 25 mM potassium phosphate pH 8.0 followed by gradient elution using 25 mM potassium phosphate buffer (pH 8.0) containing 1M NaCl. Fractions containing drUGM were collected, pooled and dialysed against 50 mM potassium phosphate pH 8.0. The sample was concentrated and brought to 30 % w/v ammonium sulfate with stirring at 4 ^{o}C. The resulting solution was filtered and applied to a pre-equilibrated HP-20 (Applied Biosystems, USA) hydrophobic interaction chromatography column with the binding buffer (50 mM potassium phosphate pH 8.0 containing 30 % ammonium sulfate). Bound proteins were eluted with decreasing gradient of ammonium sulfate in 50 mM potassium phosphate pH 8.0. Fractions containing drUGM were combined and dialyzed against 50 mM bis-tris propane pH 8.0. The purified drUGM was concentrated to 7.5 mg/ml (determined by Bradford assay) and the purity of the protein sample was judged from SDS-PAGE analysis (Figure 4-2). Small aliquots were flash-cooled using liquid nitrogen and stored at -80 ^{o}C.

3.4 Expression and purification of drUGM N372D mutant

drUGM N372D mutant was prepared by PCR-based site-directed mutagenesis method (See section 3.0 and 3.1). Expression and purification of the drUGM N372D mutant were similar as described for the drUGM WT (See section 3.3). Purified fractions were concentrated to ~ 7.5 mg/ml (as determined by Bradford assay).

3.5 Expression and purification of mtUGM WT

The gene encoding mtUGM was cloned into a pQE80 vector (in our laboratory) containing an affinity tag (six histidine residues) at the N-terminus. The construct was

transformed into *E. coli* C43 (Qiagen) cells. Transformed cells were grown in Terrific-broth medium with 100 μg/mL of ampicillin at 37 ^{o}C until the optical density reached ~ 0.6 (at 600 nm) followed by induction with IPTG (0.5 mM) for 2 hours. The cells were harvested by centrifugation at 8000 rpm for 30 minutes. The resulting cell pellet was suspended in a lysis buffer containing 50 mM of sodium phosphate, 300 mM NaCl pH 8.0, AEBSF (1mM final concentration), DNase (20 μg/ml) and lysozyme (20 μg/ml) and stirred at 4 ^{o}C for 30 minutes. Cells were further lysed by sonication and then centrifuged at 15000 rpm for 30 minutes. The supernatant was applied to a Protino Ni-IDA 1000 prepacked columns (Machery-Nagel) column (14 ml) preequilibrated with 50 mM of sodium phosphate, 300 mM NaCl, pH 8.0 (binding buffer). The column was washed with the binding buffer and the bound proteins were eluted using the elution buffer, 50 mM sodium phosphate, 300 mM NaCl and 250 mM imidazole pH 8.0. The individual fractions were analyzed for the presence of mtUGM by SDS-PAGE. Appropriate fractions were pooled and then dialyzed against 50 mM sodium phosphate buffer pH 7.0. The dialyzed sample was then concentrated and then frozen as small aliquots.

3.6 Expression and purification of kpUGM D351N mutant

kpUGM D351N mutant prepared by PCR-based site directed mutagenesis method. The primers and protocol for PCR are shown in section 3.0 and 3.1. kpUGM D351N mutant was expressed and purified in a manner similar to kpUGM WT (Chad et al, 2007). Briefly, the plasmid harboring kpUGM (in pEHISTEV vector) was transformed into *E. coli* Rosetta cells and gown in LB media at 37 ^{o}C (with kanamycin 50 μg/ml). When the optical density reached ~ 0.6 cells were induced with IPTG (0.4 mM final concentration) and incubated for a further 3-4 hours. The cell pellet was resuspended in a

lysis buffer (100 mM potassium phosphate pH 8.0, 2 mM AEBSF (4-(2-aminoethyl)-benzenesulfonyl fluoride), DNase and lysozyme 20 µg/ml) and stirred at 4 ^{o}C for 30 minutes. The cells further lysed by sonication. The suspension was clarified by centrifugation at 18000 rpm for 30 minutes. The supernatant was then treated with 10 % w/v of ammonium sulfate and stirred at 4 ^{o}C for 30 minutes. The solution was further clarified by spinning at 18000 rpm for 30 minutes. The supernatant was filtered and applied onto to a MC50 POROS column (Applied Biosystems) precharged with 50 mM $CuCl_2$ and preequilibriated with the binding buffer (50 mM of phosphate buffer, 300 mM NaCl and 0.5 mM imidiazole). The bound proteins were eluted with a gradient of 0.5 to 100 mM imidazole. Fractions were pooled, dialyzed and concentrated to 10 ml. The concentrated solution was brought to 40 % ammonium sulfate and then loaded onto a HP20 POROS hydrophobic column (Applied Biosystems) preequilibrated with the binding buffer (50 mM phosphate buffer, 40 % w/v ammonium sulfate, pH 8.0). The bound proteins were eluted with decreasing gradient of ammonium sulfate (from 40 to 0 %) and appropriate fractions were pooled, dialyzed (with 50 mM Tris-HCl pH 7.5) and concentrated to 10 mg/ml. Small aliquots were flash-frozen with liquid nitrogen and stored at -80 ^{o}C.

3.7 Kinetic characterization of drUGM WT and drUGM N372D mutant

UGM assay was performed as described (Carlson et al, 2006) with necessary modifications. The reactions were carried out in 50 mM sodium phosphate buffer pH 7.0, containing 10 nM drUGM WT or drUGM N372D mutant, UDP-Gal*f* (various concentrations) and freshly prepared sodium dithionite (20 mM final concentration). The volume of the reaction mixture was 100 µl. The time of each reaction was adjusted to

have 30-40 % conversion values (conversion of UDP-Gal*f* to UDP-Gal*p*). The reaction mixture was quenched by adding 100 µl of n-butanol. The aqueous layer was collected and injected onto a CarboPac PA1 ion-exchange column (Dionex Inc) using a Waters HPLC system (Waters 510 pump connected to Waters 717 plus Autosampler and Waters 2487 Dual λ Absorbance Detector). The sugar nucleotides were eluted isocratically with 200 mM ammonium acetate pH 7.0 and the absorbance was set to 262 nm. Analysis of the HPLC chromatogram and integration of the peaks were carried out using Waters Millennium[32] software (Version 4.0). The initial velocity was calculated from substrate concentration and % conversion. Kinetic values were determined from nonlinear regression analysis using GraphPad Prism software (GraphPad Software, San Diego California USA).

3.8 Crystallization of drUGM-ligand complexes

Initial crystallization trials to form drUGM:substrate complex were based on vapor diffusion methods by using various commercial broad screens (Hampton Research Inc). Crystal trials were carried out at 20 ^{o}C, room temperature (RT) and 4 ^{o}C. drUGM:UDP-Gal*p* complex crystals were obtained by cocrystallization method. Hits were obtained from broad screens and further optimization around the hit condition was carried out (varying pH and precipitant concentration). Crystals obtained by vapor-diffusion methods showed poor diffraction. Later changes were introduced in the cocrystallization setup and this involves chemical reduction of protein solution (by sodium dithionite) prior to substrate addition. After 20 seconds, substrate was added to the reduced protein solution with an incubation time of 5 minutes and then crystal trials were carried out. Crystals obtained in this manner showed improvement in diffraction.

Crystal conditions found in vapor diffusion methods were tried in the microbatch method (Chayen, 1997) and this showed improvement in diffraction. In addition, the effect of various cryoprotectants was studied and xylitol at 10-15 % was found to be optimal. Crystals frozen with xylitol as cryoprotectant showed better diffraction without forming ice-rings. Based on these results, the microbatch method was adopted for the crystallization of drUGM:substrate (and other ligand complexes) and are discussed below in detail.

3.8.1 Crystallization of drUGM:UDP-Gal*p* complex

Crystals of drUGM:UDP-Gal*p* ($drUGM_{ox)}$ complex were grown using the microbatch method (Chayen, 1997). Prior to crystallization, protein solution (7.5 mg/mL) was reduced with sodium dithionite (20 mM final concentration) and mixed with UDP-Gal*p* (10 mM final concentration). Equal volumes of protein solution and the crystallization solution, 0.1 M HEPES, pH 6.5, 0.2 M lithium chloride and 28 % PEG 6000 were placed in the well and covered with paraffin oil. Crystals appeared within a week and were allowed to grow for two weeks. The crystals were bright yellow color and the maximum dimensions are 0.3 mm x 0.3 mm x 0.1 mm. Crystals were harvested placing them in the cryosolution (crystallization solution with 10 % xylitol and 15 mM UDP-Gal*p*) and flash cooled with liquid nitrogen for data collection.

3.8.2 Chemical reduction of $drUGM_{ox}$ crystals

Crystals of $drUGM_{red}$ (reduced drUGM:UDP-Gal*p* complex) were obtained by chemical reduction of $drUGM_{ox}$ crystals. A drop of cryosolution (0.1 M HEPES, pH 6.5, 0.2 M lithium chloride, 28 % PEG 6000, 15 mM UDP-Gal*p* and 10 % xylitol) containing 20 mM of sodium dithionite was placed on a cover slip. Crystals of $drUGM_{ox}$ were

looped and soaked in the cryosolution and observed under a microscope. Under these conditions, $drUGM_{ox}$ crystals turned colorless (from yellow) and then immediately flash cooled with liquid nitrogen for data collection.

3.8.3 Crystallization of drUGM:UDP complex

Crystals of the drUGM:UDP complex were grown by the microbatch method using the crystallization conditions that gave drUGM-substrate crystals. Protein solution was reduced with sodium dithionite (20 mM final concentration) and mixed with UDP (10 mM final concentration). An equal volume of protein solution and crystallization solution was placed in the well and covered with paraffin oil. Crystals appeared within 5 days and allowed to grow to their maximum dimensions for 2 weeks. drUGM:UDP crystals harvested by soaking the crystals in a cryosolution containing 15 mM UDP and flash cooled with liquid nitrogen.

3.8.4 Crystallization of drUGM:UMP complex

Crystals of the drUGM:UMP complex were grown by the microbatch method using the conditions that gave drUGM:UDP crystals. Protein solution was reduced with sodium dithionite (20 mM final concentration) and mixed with UMP (10 mM final concentration). An equal volume of protein solution and crystallization solution was placed in the well and covered with paraffin oil. Crystals appeared within 10-12 days and allowed to grow to their maximum dimensions for 3 weeks. Crystals of drUGM:UMP complexes were harvested by including 15 mM of UMP in the cryosolution and flash cooled with liquid nitrogen.

3.8.5 Crystallization of drUGM:UDP-glucose complex

Crystals of the drUGM-glucose complex were grown by the microbatch method

using the conditions similar to drUGM:UDP-Gal*p* complexes. Protein solution was reduced with sodium dithionite (20 mM final concentration) and mixed with UDP-glucose (10 mM final concentration). An equal volume of protein solution and crystallization solution was placed in the well and covered with paraffin oil. Crystals appeared within 7 days and allowed to grow to their maximum dimensions for 3 weeks. Crystals were harvested by including 15 mM of UDP-glucose (UDP-Glc) in the cryosolution and flash cooled with liquid nitrogen.

3.8.6 Crystallization of drUGM:GCP complex

GCP is substrate analog of UDP-Gal*p* expected to act as an inhibitor of UGM (See section 5.0). Crystals of the drUGM:GCP complex were grown by the microbatch method using the conditions similar to drUGM:UDP-Gal*p* complexes. Protein solution was first mixed with sodium dithionite and then GCP (10 mM final concentration) was added. The drUGM:GCP complex was incubated for 5 minutes and then crystallization trials were carried out. Equal volume of protein solution and crystallization solution was placed in a well and covered with paraffin oil. Crystals appeared within 2-3 weeks. Crystals were looped and placed in a cryosolution that contains 15 mM GCP and then flash cooled with liquid nitrogen.

3.8.7 Crystallization of drUGM N372D:UDP-Gal*p* complex

Crystals of the drUGM N372D mutant were grown and harvested in a manner similar to $drUGM_{ox}$ (See section 3.8.1).

3.9 Data collection and processing

X-ray diffraction data for all drUGM complexes were collected at the Canadian Light Source (CLS), Saskatoon, Canada (08ID-1 beamline at the CLS equipped with

Marmosaic225 CCD detector). The crystal-to-detector distance was set to 250 mm with an oscillation range of 0.25^{o} and a total of 720 images (1 s exposure time for each image) were collected that covered a total oscillation range of 180^{o}. All datasets except drUGM$_{ox}$ were processed using the program D*TREK (Pflugrath, 1999). drUGM$_{ox}$ was processed using XDS (Kabsch, 2010b). Data collection and refinement statistics summarized in Table 4-3, 4-5, 4-6, 5-2 and 5-8.

3.10 Structure determination and refinement

The structure of the drUGM:substrate (drUGM$_{ox}$) complex was determined by the MR method by using the MOLREP (Vagin & Teplyakov, 2010) option from MrBUMP (Keegan & Winn, 2008) within the CCP4 package (Winn, 2003). The structure solution was found by using UGM from *M. tuberculosis* (mtUGM) as the search model (Beis et al, 2005). The initial solution had 8 copies of the monomer in the asymmetric unit (asu). This solution was fixed and a search for additional copies gave a model with 10 copies of the monomer in the asu. Refinement was carried out using REFMAC5 (Murshudov et al, 1999) and PHENIX (Adams et al, 2010; Adams et al, 2002). The structure of drUGM$_{red}$ and drUGM:UDP complexes were determined by molecular replacement using the monomer of drUGM$_{ox}$. The restraints for UDP-Gal*p* and reduced FAD were generated by using ELBOW implemented within PHENIX (Adams et al, 2010; Adams et al, 2002). Other drUGM:ligand complexes were determined by using drUGM$_{ox}$ (after removing ligands) as the starting model for refinement. Model building was performed using Coot (Emsley & Cowtan, 2004). The final structures were evaluated using the PDB validation server. Accessible area was calculated using AREAMOL program within the CCP4 package. All molecular graphics were generated using PYMOL (Delano Scientific, 2002)

3.11 Evaluation of inhibitors

Enzyme inhibition assays were performed as described with necessary modifications (Carlson et al, 2006). Reactions were monitored in the reverse direction, conversion of UDP-Gal*f* to UDP-Gal*p*, in the presence and absence of inhibitor. The extent of inhibition was expressed as % inhibition at a given concentration of the inhibitor. A time-point was chosen that gave ~ 50 % conversion of UDP-Gal*f* to UDP-Gal*p* for the evaluation of inhibitors (drUGM WT 2 minutes; drUGM N372D 3 minutes; kpUGM WT 2 minutes; mtUGM 1.30 minutes; kpUGM D351N 5 minutes). All reactions were carried out in a freshly prepared 50 mM sodium phosphate buffer, pH 7.0. Reaction vials and other solutions were degassed with argon. The enzyme (kpUGM 10 nM , drUGM WT 10 nM, drUGM N372D 50 nM, kpUGM D351N 600 nM and mtUM 87 nM) was first reduced with sodium dithionite (20 mM final concentration) and then incubated with inhibitor at 1 mM concentration for 5 minutes (for kpUGM D351N 7 minutes) and then UDP-Gal*f* was added to the reaction mixture (20 μM final concentration) and the reaction was allowed to proceed at room temperature. The reaction mixture was quenched with 100 μl of n-butanol. The aqueous layer was collected and injected into a CarboPac PA1 column (Dionex Inc), preequilibrated with 200 mM ammomuin actetate, pH 7.0. The sugar nucleotides were eluted isocratically using 200 mM ammomuin actetate, pH 7.0 with a flow-rate of 1.0 ml/minute. The two sugar-nucleotides separated with base-line resolution, and the extent of conversion was determined by integration of the two peaks (Equation 3-1).

$$\% \text{ conversion} = \frac{\text{area of UDP-Gal}p \text{ peak}}{[(\text{area of UDP-Gal}p \text{ peak}) + (\text{area of UDP-Gal}f \text{ peak})]} \times 100 \quad (3\text{-}1)$$

The percentage inhibition was calculated from the turn-over of inhibited reactions compared to the reactions with no inhibitor.

3.12 Synthesis of UDP-Gal*f*

UDP-Gal*f* was synthesized by a chemo-enzymatic method as described (Rose et al, 2008). The steps involved in the preparation of UDP-Gal*f* are discussed below.

3.12.1 Expression of GalPUT

GalPUT (Galactose-1-phosphate uridyl transferase) is an enzyme that catalyzes the transfer of UMP unit from UDP-Glc to galactofuranose-1-phosphate to form UDP-Gal*f*. GALU (Glucose-1-phosphate uridyl transferase) catalyzes the transfer of UMP from UTP to glucose-1-phosphate to form UDP-Glc.

The plasmid harboring the GalPUT gene was a generous gift from Dr. Rob Field (University of East Anglia, UK). The construct was transformed into *E. coli* BL21 cells. The cells were grown in LB media at 37 ^{o}C (100 μg/ml ampicillin) then induced with 1mM IPTG when the optical density reaches 0.6-0.8 units. Cells were allowed to grow at 30 ^{o}C for another 5 hours. The cells were then harvested by centrifugation at 8000 rpm and the cell pellet was stored at -80 ^{o}C until purification.

3.12.2 Purification and Immobilization of GalPUT

The cell pellet was thawed on ice and resuspended in 35 mL of resuspension buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0. To this, AEBSF (1 mM final concentration), lysozyme 50 μl (from 20 mg/ml stock) and DNase 50 μl (from 20 mg/ml stock) were added and stirred at 4 ^{o}C for 30 minutes. Cells were further lysed by sonication. The suspension was centrifuged at 18000 rpm for 30 minutes. The supernatant was filtered and applied onto a preequilibrated (50 mM NaH_2PO_4, 300 mM

NaCl, 10 mM imidazole, pH 8.0) Ni-NTA agarose column with a flow rate of 0.7-1.0 ml/min. The low flow rate is to ensure that there is enough time for GalPUT to interact with Ni-NTA. Unbound proteins were removed by wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0) until the absorption at 280 nM returned to baseline. Then 2 column volumes of GalPUT reaction buffer (50 mM HEPES, 10 mM $MgCl_2$, 5mM KCl, pH 8.0) was used to rinse the column and the immobilized GalPUT was stored in this buffer for the next step.

3.12.3 Coupling of UMP to Gal*f*-1-phosphate

UTP (23 mg, 40 μmol) and Gal*f*-1-phosphate (14.8 mg, 40.5 μmol) were dissolved in 0.6 ml of GalPUT reaction buffer. GALU (Glucose-1-phosphate uridyltransferase, 5U), inorganic pyrophosphatase (5U) and the immobilized resin 0.7–1.0 ml were added to the above solution. The reaction was initiated by the addition of UDP-Glc (0.15 μmol). The resulting mixture was flushed with nitrogen and the reaction stirred at room temperature. The progress of the reaction was monitored by HPLC using analytical reverse phase C18 column (4.6 x 250 mm) using 50 mM triethylammoniumacetate containing 1.5 % acetonitrile with a flow rate of 1.0 ml/min. The reaction was allowed to proceed until optimal conversion had taken place (~ 24 hours).

3.12.4 Separation of UDP-Gal*f*

The removal of immobilized GalPUT and other enzymes from the reaction mixture was accomplished by centrifugation using WMCO 10 kDa Amicon centrifugal filter devices. The final volume of the solution was made to 5 mL for the first desalting step. A sephadex G-15 (3 x 95 cm) column was used to remove salts from the reaction

buffer. Fractions containing UDP-Gal*f* were pooled together and concentrated at 25 ^{o}C (~ 10 mL). Unreacted UTP and other impurities were removed by a preparative HPLC procedure (preparative C18 reverse phase column) using 5 mM sodium phosphate buffer, pH 6.8 with a flow rate of 5 ml/min. The retention time for UDP-Gal*f* was between 11-12 minutes. Appropriate fractions were pooled and concentrated at 25 ^{o}C (~ 10 ml) and used for next desalting step. Removal of salt was achieved by using Sephadex G-15 column (3 x 95 cm). Fractions were collected and concentrated to less than 10 mL and then subjected to lyophilization.

3.13 Docking studies of inhibitors

Docking studies were performed using the program Surflex-Dock^{TM} from Tripos Inc. (Jain, 2003). Structures of inhibitor molecules were sketched and minimized using SYBYL 8.0 (Tripos, Inc). Coordinates of the drUGM:substrate complex (reduced) for docking studies (Partha et al, 2009). Hydrogen atoms were added to the protein molecule and then energy minimized. The active site was defined using UDP-Gal*p* as the reference molecule. All default parameters were used for docking studies and 10 docked conformations were requested.

CHAPTER 4

drUGM-ligand structures

4.0 UGM from *Deinococcus radiodurans* (drUGM)

Previous attempts to crystallize a UGM:substrate complex (either by soaking or cocrystallization) were unsuccessful in forming substrate bound crystals (Beis et al, 2005; Sanders et al, 2001). It has been proposed that the mobile loop located at the entrance of the substrate binding cleft may affect the formation of substrate bound crystals, either by soaking or cocrystallization (Chad et al, 2007). Due to the existing difficulties (with ecUGM, kpUGM and mtUGM) in forming UGM:substrate complex crystals, the source of the enzyme was changed. I used the UGM homolog from an alternate organism. This approach (using alternate homologs) has been proven to be successful in the crystallographic investigation of many proteins (Campbell et al, 1972; Derewenda, 2004; McPherson, 2004). UGM from *D. radiodurans* (drUGM) was used as an alternative source to determine the UGM:substrate complex structures. drUGM has sequence identities of 37 %, 42 % and 39 % to ecUGM, kpUGM and mtUGM respectively (Figure 4-1). The amino acid residues involved in FAD binding and in the active site are conserved in drUGM (Table 4-1). Based on these facts, the drUGM:substrate complex could be considered as a template for understanding the UGM:substrate structure and the active site interactions.

4.1 Purification and kinetic characterization of drUGM (WT)

The presence of an affinity tag in drUGM (six histidine residues) at the N-terminus did not help to purify the enzyme by affinity chromatography. Hence, drUGM was purified in two steps, anion-exchange chromatography followed by hydrophobic

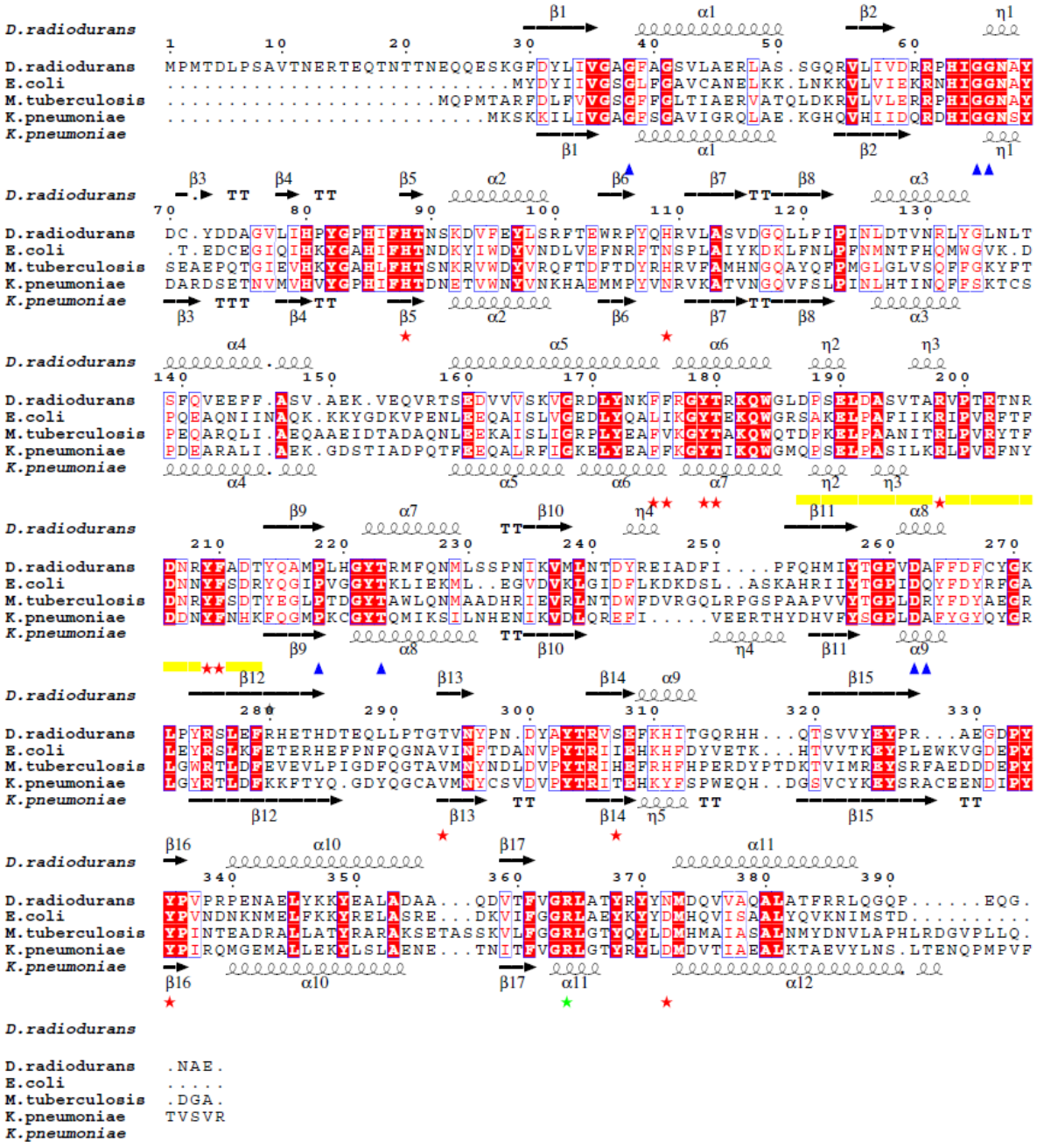


Figure 4-1 Sequence alignment of drUGM with other bacterial UGMs. The secondary structures shown above and below the sequences are based on the crystal structure of drUGM-substrate complex and kpUGM (PDB Code: 2BI7). The active site residues are shown as red asterisks, conserved residues of the FAD binding domain are highlighted in blue triangles, the mobile loop indicated by yellow bars. The conserved Arg364, which shows different orientation in the complex structure, is highlighted in green asterisk. Figure generated using *ESpript* (Gouet et al, 1999; Gouet et al, 2003).

Table 4-1 Sequence numbers for conserved active site residues of UGM from *K. pneumoniae*, *E. coli*, *M. tuberculosis* and *D. radiodurans*.

K. pneumoniae	*E. coli*	*M. tuberculosis*	*D. radiodurans*
H60	H56	H65	H88
Y155	Y151	Y161	Y179
W160	W156	W166	W184
R174	R170	R180	R198
Y185	Y181	Y191	Y209
R280	R278	R292	R305
E301	E298	E315	E325
Y314	Y311	Y328	Y335
Y349	Y346	Y366	Y370
N84	N80	H89	H109
T156	T152	T162	T180
R343	R 340	R360	R364
N270	N268	N282	N296
F151	L147	F157	F175
F152	I148	V158	F176
D351	D348	D368	N372

chromatography. The purity of the sample was judged by SDS-PAGE analysis (Figure 4-2). Steady-state kinetic analysis of drUGM (WT) was carried out to compare its kinetic parameters with other bacterial UGMs. The forward reaction, conversion of UDP-Gal*p* to UDP-Gal*f* is difficult to monitor by HPLC as the equilibrium is 93 (pyranose form):7 (furanose form) for the forward reaction. The standard UGM assay is based on monitoring the reverse reaction, i.e., conversion of UDP-Gal*f* to UDP-Gal*p* and separating the two sugar nucleotides on a HPLC column (Carlson et al, 2006; Zhang & Liu 2001). The separation of sugar nucleotides were done using a CarboPac PA1 column (Carlson et al, 2006), although, C18 reverse phase columns were used by others (Caravano et al, 2003; Zhang & Liu 2001). The separation of sugar nucleotides was better with the CarboPac PA1 column as compared to C18 reverse phase columns. All reactions were carried out at room temperature in the presence of sodium dithionite (20 mM). The standard Michaelis-Menten saturation curve for each concentration of substrate is shown

in Figure 4-3. The kinetic parameters of drUGM WT were comparable to other bacterial UGMs (Table 4-2).

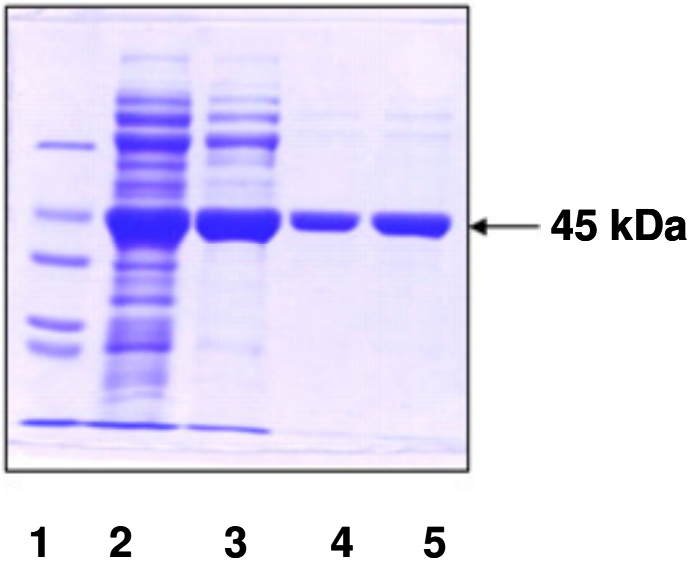


Figure 4-2 SDS-PAGE analysis of purified fractions of drUGM. Lane 1 low molecular weight marker, lane 2 supernatant after cell lysis, lane 3 fractions after anion-exchange chromatography, lane 5 (10 µl) and lane 6 (15 µl) fractions after hydrophobic chromatography.

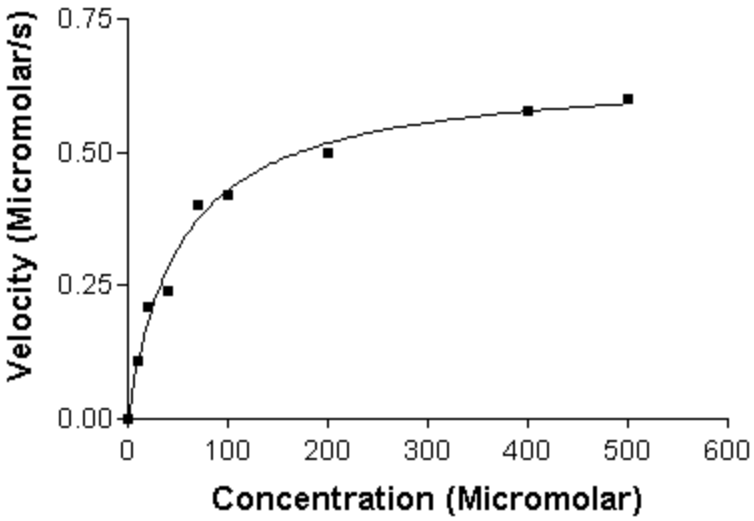


Figure 4-3 Saturation curve for drUGM for each concentration of the substrate.

Table 4-2 Kinetic data for drUGM and other bacterial UGMs.

Enzyme	$k_{cat}(s^{-1})$	K_m (μM)	k_{cat}/K_m (μM^{-1} s^{-1})
ecUGM[1]	22	27	0.81
kpUGM[2,3]	5.0 ± 0.6	45 ± 6.0	0.11
drUGM[4]	66 ± 2.4	55 ± 7.0	1.18

[1]The authors of this paper (Zhang & Liu, 2000) did not report the errors for the kinetic parameters.
[2](Chad et al, 2007).
[3]A K_m value of 16 μM and k_{cat} of 2210 min^{-1} reported for kpUGM with 5-10 % error in the data (Errey et al, 2009).
[4]Kinetic values reported in this thesis.

4.2 Structure of oxidized drUGM:UDP-Gal*p* complex ($drUGM_{ox}$)

Crystals of $drUGM_{ox}$ were grown by cocrystallization using the microbatch method (Karunan Partha et al, 2009). The $drUGM_{ox}$ complex crystallized in the $P2_12_12_1$ space group and the crystals diffracted to 2.4 Å. Data collection and refinement statistics are shown in Table 4-3. The structure solution was obtained by molecular replacement (MR) by using mtUGM (PDB code: 1V0J) as the search model (Beis et al, 2005). The MOLREP program (Vagin & Teplyakov, 2000) implemented within MrBUMP (Keegan & Winn, 2008) was used for the MR search. The crystals of $drUGM_{ox}$ contain 5 homodimers in the asymmetric unit (asu), all identical. The electron density for the sugar (galactose) moiety in $drUGM_{ox}$ was weak; therefore, NCS averaging over the 10 active sites was carried out to generate an averaged map for modeling the galactose portion of the substrate in the $drUGM_{ox}$.

UGM is a homodimer and belongs to the α/β class of protein. Each monomer of UGM can be divided into three distinct domains (Figure 4-4). Domain 1 is the binding site for FAD, containing the αβα Rossmann fold. Domain 2 is a five-helix bundle and

domain 3 is a six-stranded anti-parallel β-sheet. The substrate binding site is located in a cleft adjacent to the isoalloxazine ring of FAD. Two flexible loops (loop 1 and loop 2)

Table 4-3 Data collection and refinement statistics.

	$drUGM_{ox}$	$drUGM_{red}$	drUGM:UDP
Data collection			
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions			
a, *b*, *c* (Å)	134.0, 176.6, 221.6	132.8, 174.6, 218.1	134.05, 176.87, 222.9
	90.00, 90.00, 90.00	90.00. 90.00, 90.00	90.00, 90.00, 90.00
Resolution (Å)	2.36 (2.42-2.36)	2.50 (2.59-2.50)	2.55 (2.649-2.55)
R_{sym}	11.6 (68.9)	18.0 (61.0)	17.2 (57.8)
$I / \sigma I$	13.0 (3.0)	5.5 (1.9)	7.2 (1.0)
Completeness (%)	95.8 (87.4)	96.3 (93.9)	98.9 (93.1)
Redundancy	7.2	4.8	7.9
Refinement			
Resolution (Å)	20.0-2.36	40.0-2.50	40.0-2.55
No. reflections	204925	168600	204114
R_{work} / R_{free}	0.17/0.22	0.21/0.27	0.18/0.24
No. atoms			
Protein	30998	30009	30748
Ligand (FAD and UDP-Gal*p*)	880	880	780
Water (molecules)	1241	556	1096
B-factors			
Protein	36.0	53.8	44.6
Ligand	39.3 (UDP-Gal*p*)	51.9 (UDP-Gal*p*)	41.7 (UDP)
	37.9 (FAD)	53.2 (FAD)	44.5 (FAD)
Water	41.5	47.8	43.4
r.m.s. deviations			
Bond lengths (Å)	0.008	0.009	0.009
Bond angles (°)	1.2	1.2	1.2

*values in the parenthesis are for the highest resolution shell

are located at the entrance of this cleft and are believed to move upon substrate binding (Sanders et al, 2001).

The overall structure of $drUGM_{ox}$ is similar to the other bacterial UGM structures, except for the mobile loop conformation (loop 1 and 2) (Figure 4-4). To date, efforts

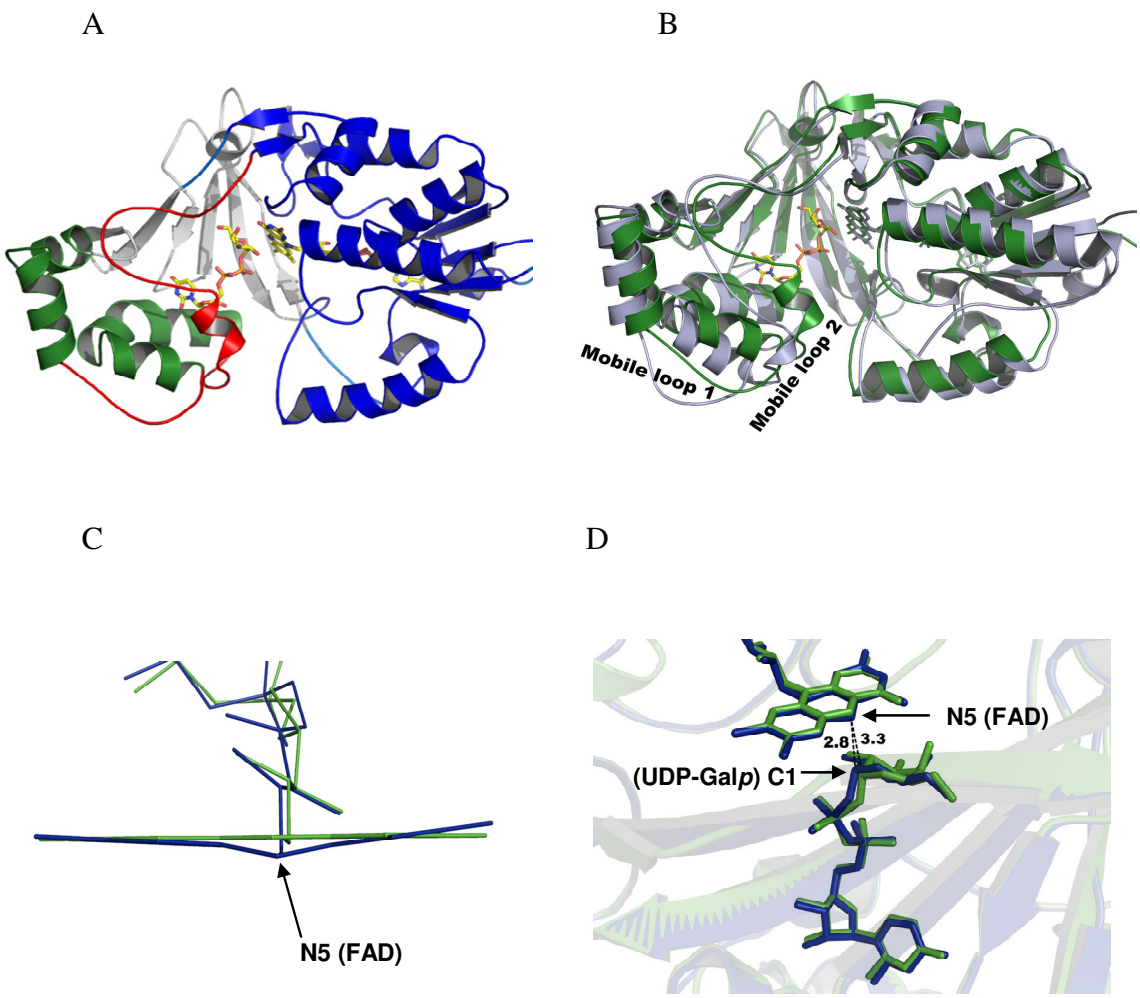


Figure 4-4 Crystal structure of drUGM-substrate complex. **(A)** Ribbon diagram representation of a monomer of $drUGM_{ox}$. Domain 1 (FAD binding domain), 2 and 3 are colored blue, green and gray respectively while the flexible loop 1 and 2 are colored red. The substrate (UDP-Gal*p*) and FAD are rendered as sticks. **(B)** Superposition of kpUGM (cyan) with $drUGM_{ox}$ (green). Note the closed conformation of the loop 2 in $drUGM_{ox}$. **(C)** Comparison of the flavin conformation in $drUGM_{ox}$ (green) and $drUGM_{red}$ (blue). **(D)** Overlay of $drUGM_{ox}$ (green) and $drUGM_{red}$ (blue). The N5 of FAD in $drUGM_{red}$ is close to the C1 of the galactose.

to determine the structure of unliganded of drUGM were unsuccessful. But, it is expected that the structure of unliganded drUGM will be similar to the other unliganded bacterial UGMs. In order to compare the drUGM:substrate complex structure with the substrate

free UGM structure, unliganded kpUGM was used (Beis et al, 2005). The sequence identity between kpUGM and drUGM is 42 %. Monomer A from $drUGM_{ox}$ could be superimposed with monomer A of kpUGM with a rmsd value of 1.4 Å for 350 equivalent C_α atoms including the two mobile loops (Figure 4-4). The individual domains have rmsd values of 0.9 Å, 1.9 Å and 1.4 Å for domain 1, 2 and 3 respectively. Domain 2 and 3 contains most of the active site residues, while domain 1 is primarily involved in FAD binding.

4.3 Structure of reduced drUGM:UDP-Gal*p* complex ($drUGM_{red}$)

Crystals of $drUGM_{ox}$ were reduced chemically with sodium dithionite to obtain $drUGM_{red}$ crystals. After brief exposure to sodium dithionite, the crystals turned from yellow to colorless, indicating that flavin was reduced. The structure of $drUGM_{red}$ was solved by molecular replacement with MOLREP (Vagin & Teplyakov, 2000) using monomer A of $drUGM_{ox}$. $drUGM_{red}$ crystals diffracted to 2.5 Å. Data collection and refinement statistics are shown in Table 4-3. The overall structure of $drUGM_{ox}$ and $drUGM_{red}$ are similar (including the two flexible loops), with a rmsd value of 0.72 Å over all equivalent C_α atoms (Partha et al, 2009). Notable differences between $drUGM_{ox}$ and $drUGM_{red}$ are the electron density for the sugar moiety and the FAD conformation (Figure 4-4). The electron density for the sugar moiety is well defined in $drUGM_{red}$ compared with $drUGM_{ox}$, suggesting stabilization of the sugar conformation in reduced UGM (Figure 4-5). The isoalloxazine ring of FAD in the $drUGM_{red}$ structure showed a bent conformation with the N5 of $FADH^-$ pointing towards the sugar moiety of the substrate. In $drUGM_{red}$ the density of FAD in some monomomers corresponded to oxidized FAD. It is possible that there is a mixed population of oxidized and reduced

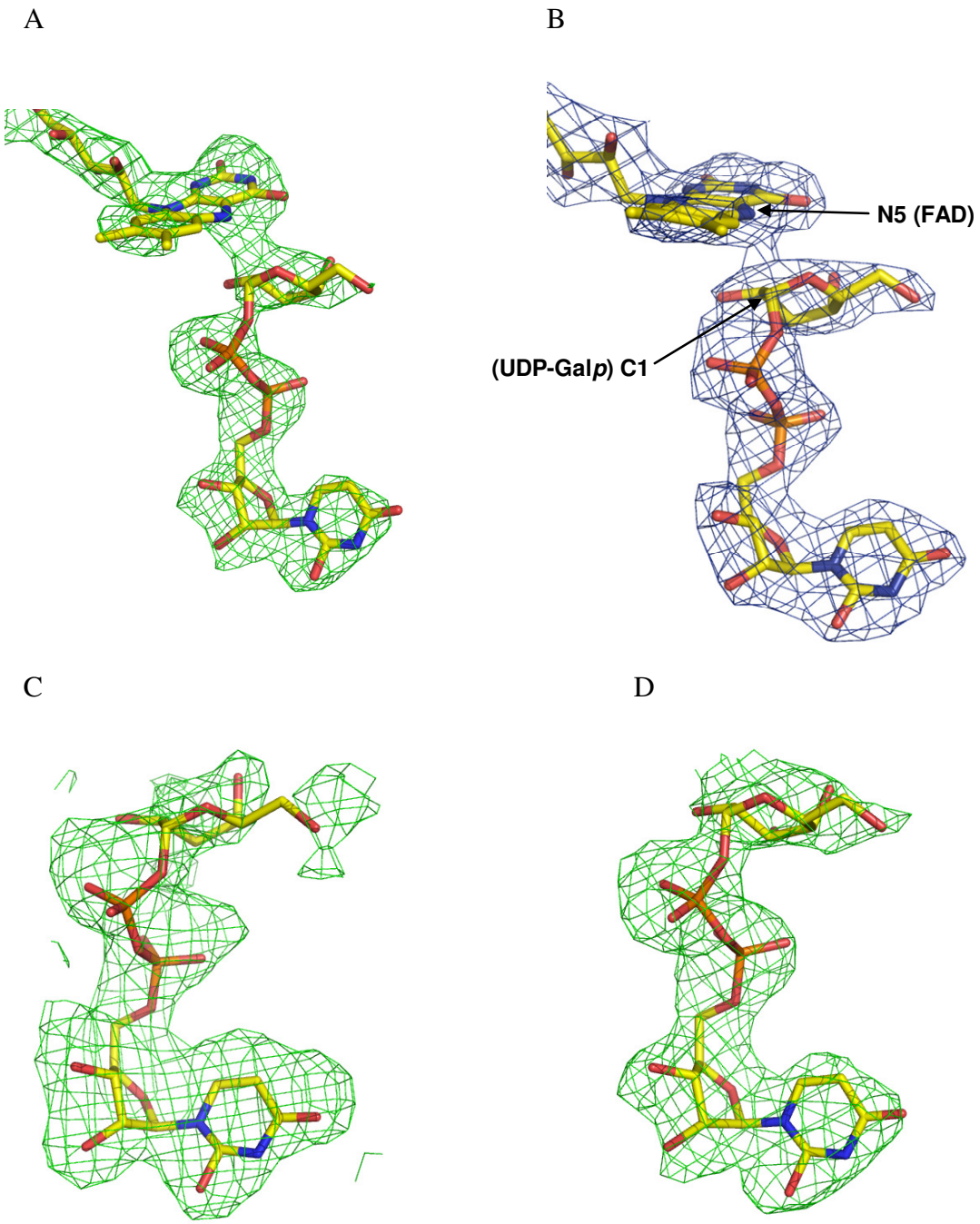


Figure 4-5 Electron density maps of UDP-Gal*p* A) Omit map of drUGM$_{red}$ contoured to 2.0 σ calculated after simulated annealing without FAD and UDP-Gal*p*. FAD and UDP-Gal*p* are modeled into the density for further refinement. B) The refined $2F_o$-F_c map of drUGM$_{red}$ contoured to 1.0 σ. Note the residual density between the N5 of FAD and the sugar moiety of the substrate. C) NCS averaged difference omit map (contoured at 4.0 σ) of the substrate in drUGM$_{ox}$. D) Difference map of the substrate in drUGM$_{red}$ (chain D) contoured to 2.0 σ. The density for the sugar moiety is well defined in D (drUGM$_{red}$) compared to C (drUGM$_{ox}$).

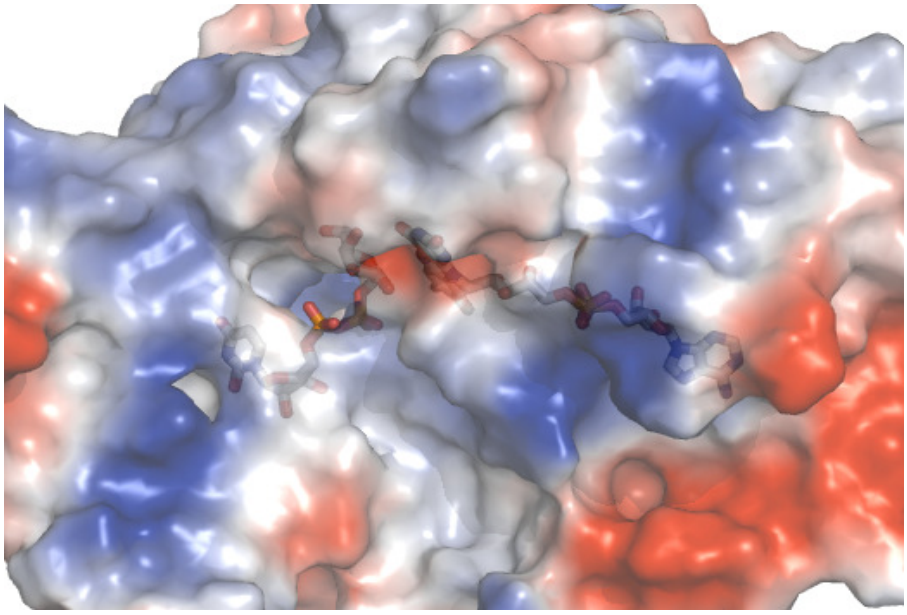
FAD in the crystals.The density for reduced FAD was observed in chain B, D and G. The distance between N5 of FAD to the anomeric carbon of galactose in $drUGM_{ox}$ and $drUGM_{red}$ was found to be 3.3 Å and 2.8 Å respectively (Figure 4-4). The residual density between the reduced FAD and the sugar moiety was observed only in chain D and is suggestive of bond formation between them (Figure 4-5). Previously, a tryptophan fluorescence assay showed that reduced UGM has a 3-fold greater affinity for substrate than oxidized UGM (66 μM vs. 220 μM) (Yao et al, 2009). The structural observations for $drUGM_{red}$, including the stabilized sugar conformation and possible bond formation are consistent with these previous studies.

4.4 Substrate binding model

The binding mode of the substrate and its active site interactions are shown in Figures 4-6, 4-7, 4-8 and 4-9. The substrate is buried in the active site with only 6.0 $Å^2$ (0.85% of the total surface area) exposed to the bulk solvent (Figure 4-6A) and is bound in a folded U-shaped conformation (Partha et al, 2009).

The binding mode of UDP-Gal*p* in drUGM is remarkably different from the fully extended or folded conformations of UDP-Gal*p* (or fluoro analog of UDP-Gal*p*) observed in the structures of other enzymes (Figure 4-6B), such as α-1,4-galactosyltransferases (*from Neisseria menintigitidis*) and UDP-galactose-4-epimerase (Persson et al, 2001; Thoden & Holden, 1998). Based on the substrate (UDP-Gal*p*) structure, the active site of UGM can be divided into three different regions, namely, the uridine binding pocket, the phosphate binding region and the sugar binding cleft.

A



B

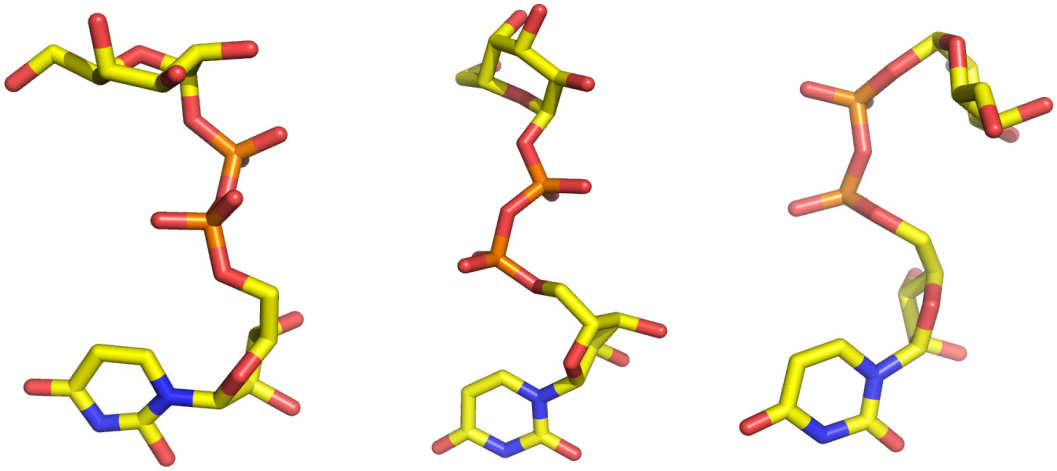


Figure 4-6 Binding mode of UDP-Gal*p*. **A)** Surface representation of drUGM$_{ox}$. Substrate and FAD represented as sticks. The surface (red, negatively charged region; blue, positively charged region) is rendered transparent to highlight the buried substrate. **B)** Comparison of UDP-Gal*p* binding conformation in different crystal structures. Conformation in drUGM$_{ox}$ **1,** UDP-galactose epimerase **2** (PDB code 1A9Z) and in galactosyltransferase **3** (PDB code 1G9R).

4.4.1 Uridine binding pocket

In drUGM:substrate complex structures, residues in the uridine binding pocket (Figure 4-7) showed very little movement when compared with the unliganded UGM (Partha et al, 2009; Sanders et al, 2001). Interactions in this region are dominated by a network of hydrogen bonds between the uridine moiety and the side chain of amino acid residues. Based on modeling and kinetic studies, an active site tryptophan (Trp160 in kpUGM) was identified and assigned a critical role in substrate binding and activity (Chad et al, 2007). Modeling suggest that Trp160 can form π-stacking intractions with uracil ring of the substrate (Chad et al, 2007; Sanders et al, 2001; Yuan et al, 2008). In

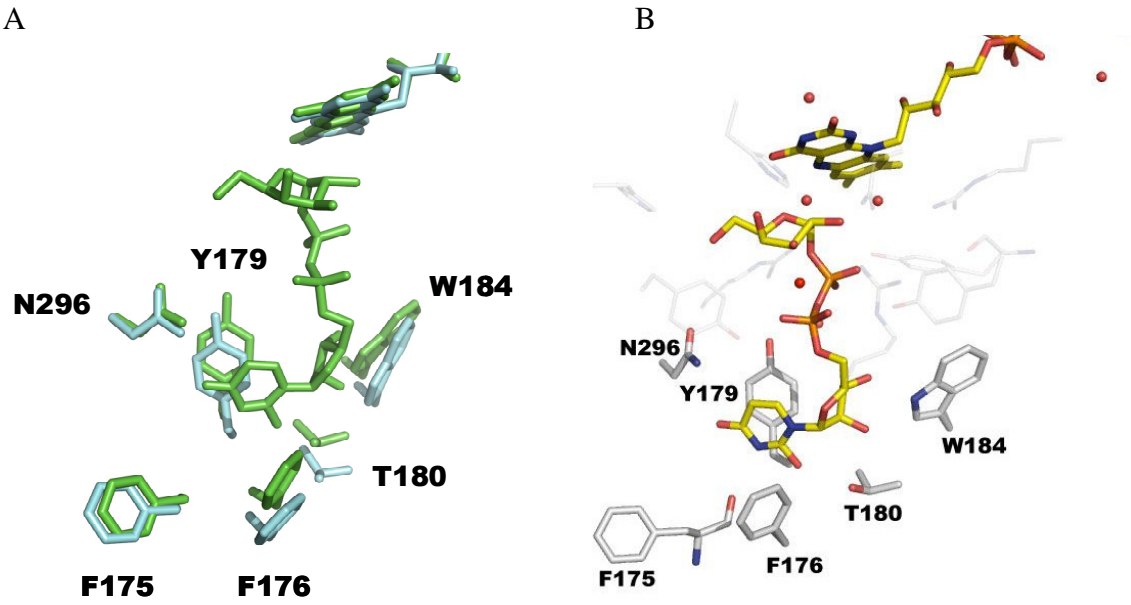


Figure 4-7 Active site interactions of UDP-Gal*p* at the uridine binding pocket. **A)** Overlay of unliganded kpUGM and $drUGM_{ox}$ to highlight the orientation of residues in uridine binding region. **B)** Active site interactions at the uridine binding pocket. Water molecules represented as red spheres.

drUGM:substrate complex structures the uracil ring is stacked between Tyr179 and Phe176 (3.9 Å from the centre of the aromatic ring of Tyr 179 to the centre of uracil ring and 4.8 Å for the corresponding distance between Phe 176 and the uracil). The indole NH

of Trp184 (Trp160 in kpUGM) form hydrogen bonds with the C3' (2.8 Å) and C2' (3.3 Å) hydroxyl groups of ribose. This is still consistent with a critical role for Trp184 in substrate binding. A similar binding mode for the uridine moiety has been suggested in other modeling studies (Bleile, 2008) and is also seen in the kpUGM:UDP-Gal*p* complex structure (See section 4.6) (Gruber et al, 2009a; Gruber et al, 2009b). Additional hydrogen bonds are formed between O2, O4 and N3 of uracil with Thr180, Asn296 and Phe175 respectively. The C2' hydroxyl group of ribose also forms a hydrogen bond with the Oγ1 of the conserved Thr180. The only observed conformational change in the uridine binding pocket is the orientation of Gln183 side chain (from domain 2). In $drUGM_{ox}$ and $drUGM_{red}$ the side chain of Gln183 is oriented away from the active site. In unliganded UGM structures the corresponding Gln is directed towards the active site. This conformational change is likely to avoid steric clashes between the substrate (ribose moiety) and the side chain of Gln183.

4.4.2 Phosphate binding region

The two phosphate groups (α and β) of the substrate form salt-bridges with the highly conserved arginines, Arg198 and Arg305 (Figure 4-8). Notably, Arg198 located in the mobile loop 2 stabilizes the α-phosphate of the substrate and the β-phosphate group is stabilized by Arg305. Mutation of either residue (mutants of kpUGM, R174A and R280A) abolish enzyme activity (Chad et al, 2007). The two phosphate groups of the substrate are further stabilized by hydrogen bond formation with conserved tyrosines (Tyr209, Tyr370 and Tyr335). These residues were shown to be important for substrate binding (Chad et al, 2007; Sanders et al, 2001). Three water molecules located in the

active site function as bridges to form hydrogen bonds between the two phosphate groups and the side chains of active site residues.

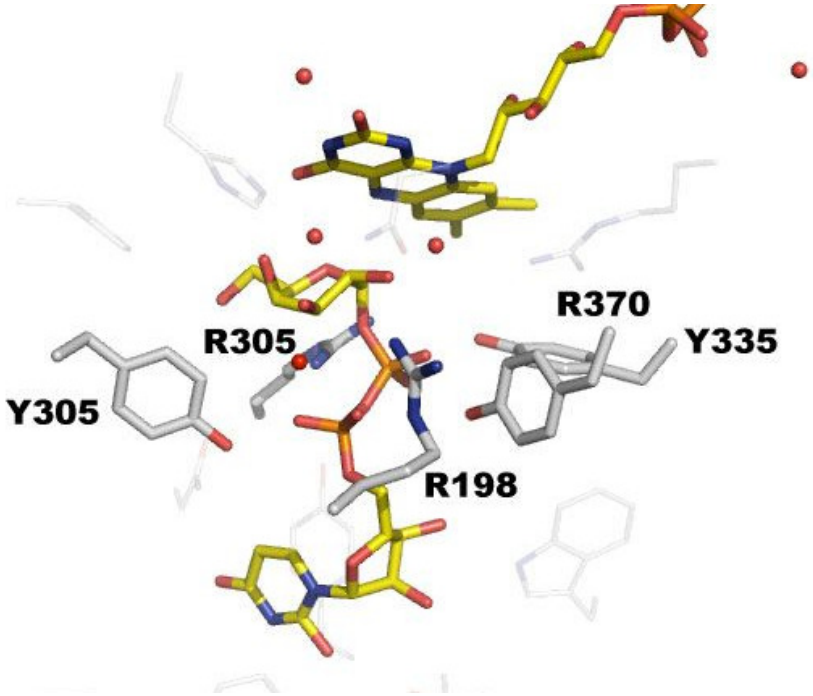


Figure 4-8 Active site interactions of UDP-Gal*p* at the phosphate binding region.

4.4.3 Sugar binding cleft

The sugar (galactose) moiety of the substrate occupies a pocket adjacent to the isoalloxazine ring of FAD (Figure 4-9). It is stabilized primarily through water mediated hydrogen bonds with the side chains of His88, His109, Arg364, and Asn372. Tyr371 forms water mediated hydrogen bonds with the sugar moiety through its backbone carbonyl oxygen atom. The C2 and C3 hydroxyl groups of galactose are involved in water mediated hydrogen bonding with His88, Asn372, Arg364 and Tyr371. C4 and C6 hydroxyl groups form direct hydrogen bonds with O4 of FAD and NE1 of His109 respectively. The hydrogen bond between O4 of FAD and the C4 hydroxyl group of galactose plays an important role in substrate recognition and activation of the C4

hydroxyl group. The C4 hydroxyl of galactose (axial orientation) can form strong hydrogen bond interactions with O4 of FAD. This interaction may be weak if the C4 hydroxyl group is equatorial (in UDP-Glc). This may account for the specificity of UGM for UDP-Gal*p* over UDP-Glc (See section 4.10). This hydrogen bond between the O4 of FAD and the C4 hydroxyl group may assist in the removal of a proton from the C4

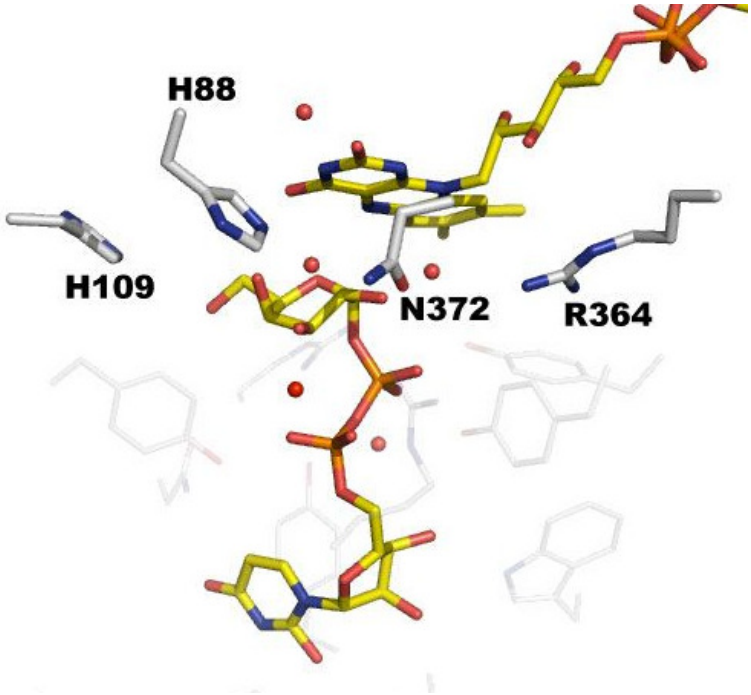


Figure 4-9 Active site interactions of UDP-Gal*p* at the sugar binding cleft.

hydroxyl group and increase its nucleophilicity to initiate the ring closure step (Figures 1-4 and 1-5). The anomeric carbon of galactose in $drUGM_{red}$ is located at a distance of 2.8 Å from N5 of $FADH^-$. The proximity of the anomeric carbon to N5 of $FADH^-$ is favorable to form a FAD-galactosyl adduct and is consistent with the proposed role for FAD in the reaction (Huang et al, 2003; Soltero-Higgin et al, 2004). Overall, the UDP portion of the substrate is stabilized by interactions with active site amino acid residues, while the sugar moiety interacts mainly through water mediated hydrogen bonds. This

pattern of interaction is consistent with the STD-NMR studies of UGM-substrate interactions, where the protons of the UDP moiety showed the largest STD effects (Yuan et al, 2005). The binding mode of substrate is also consistent with the recent report on 2-3- and 6-fluoro-substituted UDP-Gal*p* analogs that were found to be substrates for kpUGM (Errey et al, 2009). It has been shown that none of the 2, 3 and 6-hydroxyl groups of galactose in UDP-Gal*p* are essential for substrate binding and turnover. In drUGM:UDP-Gal*p* complex structures, the C2 and C3 hydroxyl groups of galactose are not involved in direct contact with any amino acid residues and they form water mediated hydrogen bonds. The C6 hydroxyl group forms a hydrogen bond with His109 (3.3 Å). These fluoro analogs are capable of forming hydrogen bonds and were able to act as substrates for UGM. Analogs with the C2 hydroxyl in axial orientation, namely, UDP-talose and UDP-2-deoxy-2-fluoro-talose (2-fluoro in axial orientation) are not substrates of UGM (Errey et al, 2009). The drUGM:substrate complex structures provide a molecular basis for the inability of UDP-talose or UDP-2-deoxy-2-fluoro-talose to act as substrate or inhibitor for UGM. The substrate binding mode in drUGM suggests that C2 axial substitutions would result in severe steric clashes with FAD and thus any efforts to prepare C2 axial substituent may render the compound inactive.

4.5 Rearrangement of flexible loop

Based on the structural, modeling and biochemical studies on UGM it has been proposed that the two mobile loops undergo rearrangement upon substrate binding (Chad et al, 2007; Sanders et al, 2001). Most importantly, mobile loop 2 undergoes significant rearrangement upon substrate binding from an open to closed conformation. In drUGM:substrate complex structures the mobile loops (loop 1: 149-157 and loop 2: 185-

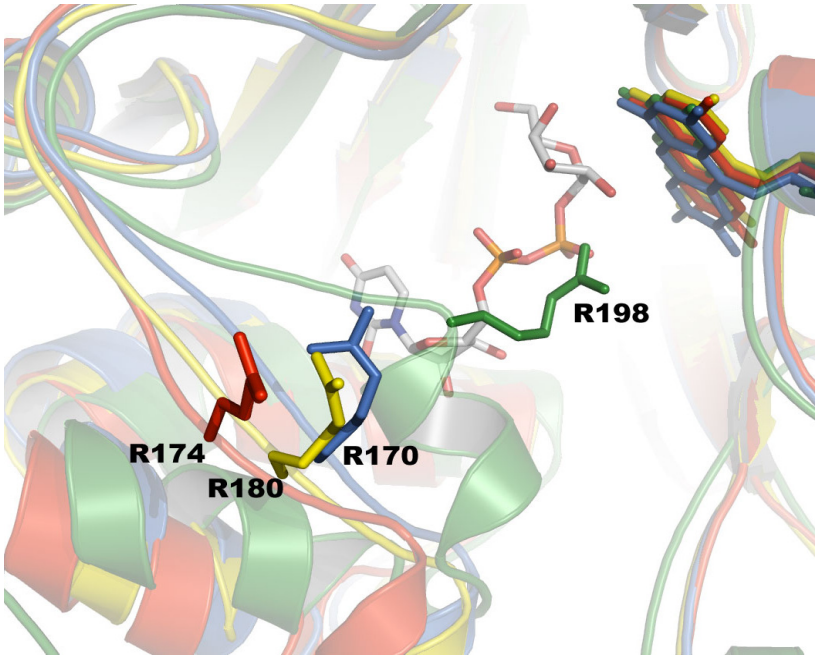
215) have moved significantly towards the substrate binding pocket and exist in a closed conformation as compared to the open conformation in other bacterial unliganded UGMs (Figure 4-10). The distance between the C_α of Arg198 (in drUGM:substrate complex) and the corresponding arginine in unliganded UGMs is ≥ 7.0 Å (Table 4-4). Arg198 of the mobile loop 2 stabilizes the closed conformation by forming salt-bridges with the α-phosphate group of the substrate and also by cation-π interaction with Tyr370. In ecUGM, monomer A and B exist in a closed and open conformation respectively (Sanders et al, 2001). Overlay of monomer A of ecUGM (closed conformation) with the substrate complex structure is shown in Figure 4-10. The conserved arginine (Arg170 in ecUGM) of the flexible loop is pointing away from the active site, whereas in drUGM:substrate complex structures the corresponding arginine (Arg198) orients towards the substrate and stabilizes its α-phosphate group (2.8 Å from Nε of Arg198 to the oxygen of the α-phosphate group). drUGM:substrate complex structures reveal the structural and functional role of the conserved arginine (Arg198) of the flexible loop and it is consistent with the other biochemical studies (Chad et al, 2007; Yao et al, 2009; Yuan et al, 2008).

Table 4-4 Distance between the C_α of Arg198 (drUGM:substrate complex) and the corresponding arginine in other bacterial UGMs.

Enzyme	Residue Number	Distance (Å)
ecUGM (monomer B)	R170	7.0
ecUGM (monomer A)	R 170	2.2*
kpGM	R174	10.3
mtUGM	R180	8.5

*Athough monomer A of ecUGM is in a closed conformation, the side chain of Arg170 is pointing away from the substrate binding cleft (Figure 4-8B).

A



B

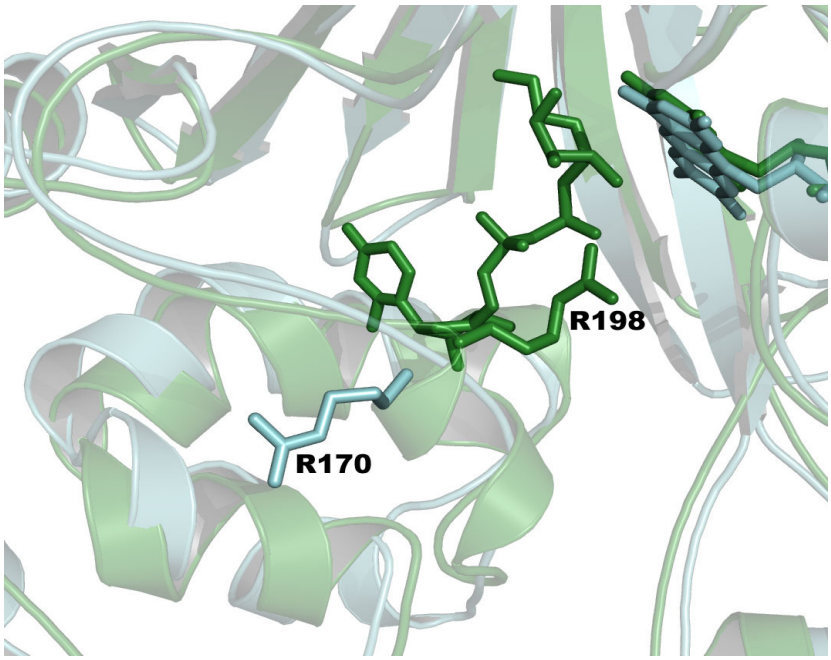


Figure 4-10 Conformational changes in drUGM:substrate complex structures **(A)** Comparison of flexible loop conformation of drUGM:substrate complex with unliganed UGMs. $drUGM_{ox}$ (green), kpUGM (red), ecUGM, monomer B (blue) and mtUGM (yellow). **(B)** Overlay of structures of $drUGM_{ox}$ (green) and closed conformer of monomer A of ecUGM (Cyan). Arg170 (in ecUGM) of the flexible loop is facing away from the substrate binding site while the corresponding arginine of $drUGM_{ox}$ (Arg198) point towards the substrate binding pocket and stabilize the α-phosphate group of UDP-Gal*p*.

4.6 Conformational changes in FAD binding domain

In $drUGM_{ox}$ and $drUGM_{red}$, the side chains of Arg364 and Tyr370 showed significant differences in their orientation when compared with the unliganded UGMs (Figure 4-11). In the unliganded UGM structures the corresponding side chains of arginine and tyrosine interact with the phosphate group and N10 of FAD respectively (Arg343 and Tyr349 in kpUGM). In the drUGM:substrate structures the side chain of Arg364 is oriented towards the active site and forms a water-mediated hydrogen bond with the C2 hydroxyl group of galactose (Partha et al, 2009). The side chain of Tyr370 (in drUGM:substrate complex) is tilted at an angle of ~ 45º (as compared to Tyr349 of kpUGM) and interacts with the α-phosphate group of the substrate. The side chain of Tyr370 forms hydrogen bonds with the phosphate group of the substrate and the aromatic

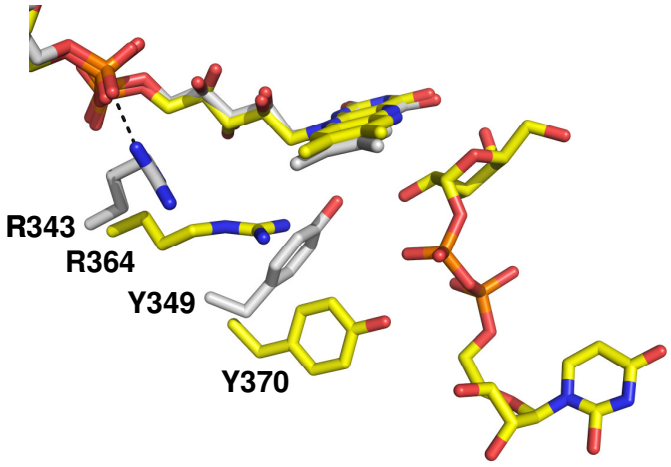


Figure 4-11 Overlay of structures of $drUGM_{ox}$ and unliganded kpUGM. Note the orientation of the Arg364 and Tyr370 of $drUGM_{ox}$ and the equivalent residues in kpUGM (Arg343 and Tyr349).

ring is stacked against Arg198 (cation-π interactions). It is possible that these conformational changes lock the entrance of the pocket (together with the Arg198) and act as a gateway for the entrance and release of the substrate and product.

4.7 Comparsion of drUGM:UDP-Gal*p* and kpUGM:UDP-Gal*p* complexes

Structures of drUGM:substrate complexes were the first substrate bound structure for any UGM to be deposited in the PDB (3HDQ and 3HDY). Later, structures of oxidized and reduced kpUGM:substrate complexes were reported after we submitted our results for publication (Gruber et al, 2009b; Partha et al, 2009). drUGM and kpUGM share 42 % sequence identity and the active site residues are conserved between them. The overall structure, binding mode and active site interactions between the two complexes are expected to be similar. Indeed, the overall structures of the two complexes were similar, but major differences between oxidized substrate complexes ($drUGM_{ox}$ vs $kpUGM_{ox}$) and minor differences between reduced substrate complexes were observed ($drUGM_{red}$ vs $kpUGM_{red}$). In $kpUGM_{ox}$, the mobile loops in monomer A and B don't exist in an open conformation, while in $drUGM_{ox}$ a fully closed conformation observed in all the chains. (Figure 4-12). The binding mode of UDP-Gal*p* in $kpUGM_{ox}$ is significantly different in $kpUGM_{ox}$ when compared with $drUGM_{ox}$. For example, the stabilization of the phosphate groups by the two conserved arginines (Arg174 and Arg280) is completely missing in $kpUGM_{ox}$, whereas in $drUGM_{ox}$ the two conserved arginines (Arg198 and Arg305) are in close contact with the phosphate groups of the substrate. The location of the sugar moiety and its binding mode is also significantly different in $kpUGM_{ox}$ when compared with $drUGM_{ox}$. The sugar moiety is located at the entrance of the binding cleft in $kpUGM_{ox}$. In $drUGM_{ox}$ the sugar moiety is in close proximity to the isoalloxazine ring of FAD. The anomeric carbon of the sugar moiety is located at a distance of 8.0 Å in both the monomers for $kpUGM_{ox}$ and the corresponding

distance in $drUGM_{ox}$ is 3.3 Å which is consistent with the predictions from modeling studies.

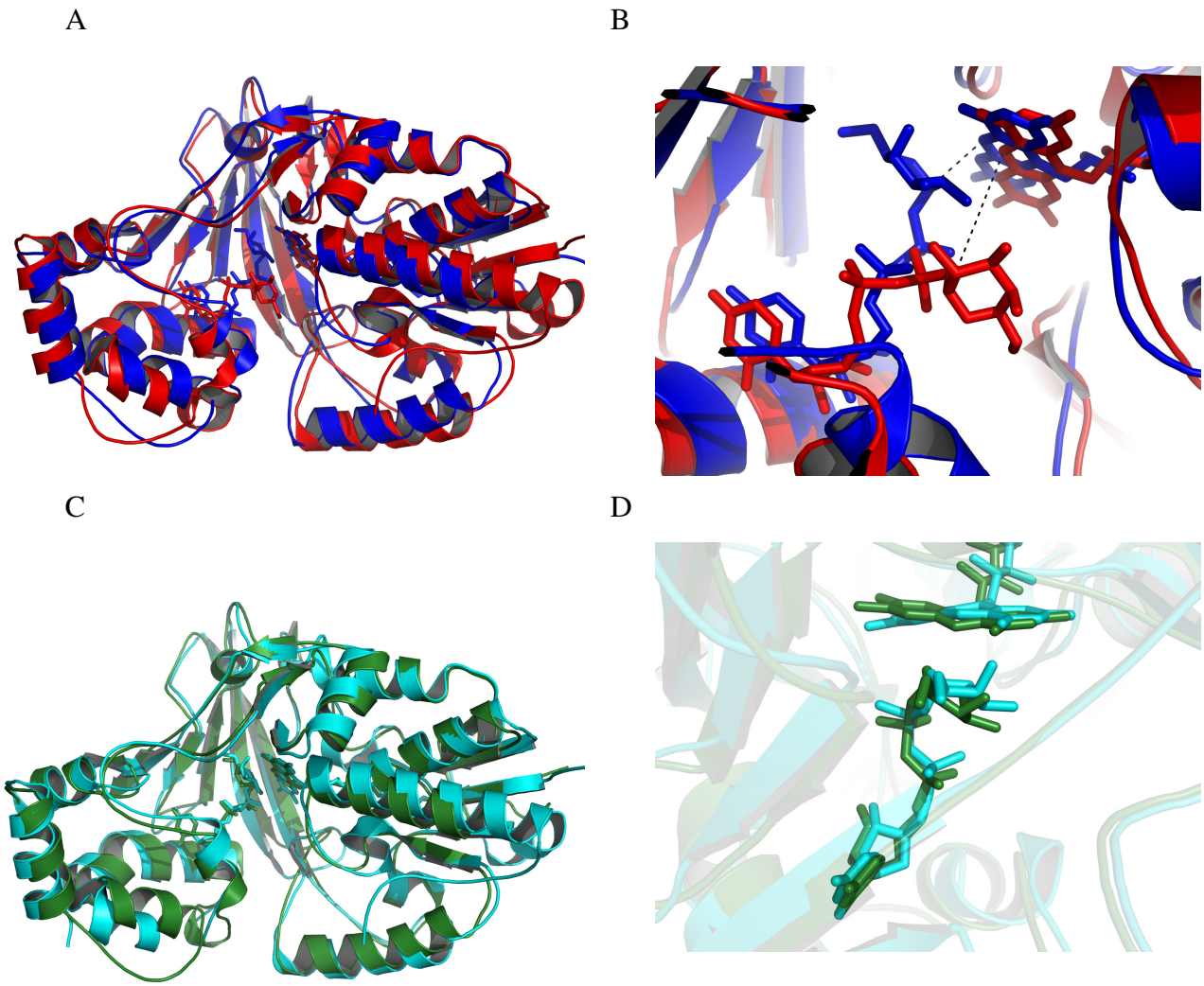


Figure 4-12 Comparison of drUGM and kpUGM substrate complex structures. **A)** Over lay of $drUGM_{ox}$ (blue) and $kpUGM_{ox}$ (red). Note the fully closed conformation of the mobile loop 1and 2 in $drUGM_{ox}$. **B)** Overaly of $drUGM_{ox}$ and $kpUGM_{ox}$ to highlight the binding mode differences of UDP-Gal*p*. The sugar moiety is orieted close to FAD in $drUGM_{ox}$, while in $kpUGM_{ox}$ it is away from the FAD. Also, note the distance between the anomeric carbon of galactose in $drUGM_{ox}$ and $kpUGM_{ox}$. **C)** Overlay of $drUGM_{red}$ (Green) and $kpUGM_{red}$ (cyan). The mobile loop 2 adopt similar conformation both the complexes, however, mobile loop 1 is missing in the kpUGMred (poor density in this region). The binding mode of UDP-Gal*p* is similar in both the complexes. **D)** Overlay of $drUGM_{red}$ (Green) and $kpUGM_{red}$ (cyan) to highlight the puckering of the isoalloxazine ring of FAD. In $kpUGM_{red}$ the N5 is pointing towards the protein, while, in $drUGM_{red}$ the N5 atom of FAD is puckered towards the substrate.

The binding mode of UDP-Gal*p* in $drUGM_{red}$ and $kpUGM_{red}$ (monomer B) is similar and the interaction patterns of the substrate were identical in both the complexes (Figure 4-12). In $kpUGM_{red}$ monomer A has UDP (not UDP-Gal*p*) in the active site and its binding mode is significantly different from the UDP portion of the substrate in monomer B. The mobile loop of $kpUGM_{red}$ (monomer A) is not in a fully closed conformation, when compared with the mobile loop of monomer B. The phosphate groups of UDP in monomer A are not stabilized by the two conserved arginines (Arg174 and Arg280). In $drUGM_{red}$, a fully closed conformation is observed in all the chains and the two conserved arginines (Arg198 and Arg305) stabilize the phosphate groups of UDP-Gal*p*. Also, differences in FAD conformation were observed between $kpUGM_{red}$ and $drUGM_{red.}$ In both complexes the isoalloxazine ring of reduced FAD exist in a bent conformation, but the orientation of N5 of isoalloxazine ring was found to be different between the two complexes. In $kpUGM_{red}$, N5 is pointing towards backbone carbonyl oxygen of Pro59 and the distance between N5 of FAD to the anomeric carbon is 3.6 Å (monomer B). In $drUGM_{red}$, N5 of reduced FAD is orient towards the substrate (*re*-face) and distance between the N5 of reduced FAD and the anomeric carbon is 2.8 Å. Based on these differences, $drUGM_{red}$ can be considered as a closer approximation for reduced UGM:substrate complex.

4.8 Structure of drUGM:UDP complex

Crystals of drUGM:UDP complexes were grown by the microbatch method using conditions similar to $drUGM_{ox}$. drUGM:UDP complex crystallized in the $P2_12_12_1$ space group and the unit cell dimensions were similar to drUGM:substrate complex structures. Crystals of drUGM:UDP complex diffracted to 2.5 Å and the structure was determined

by the MR method. drUGM$_{ox}$ was used as the search model for MR using MOLREP (Vagin & Teplyakov, 2000). Data collection and refinement statistics are shown in Table 4-3. The asu of the drUGM:UDP crystal contains 5 homodimers (Partha et al, 2009).

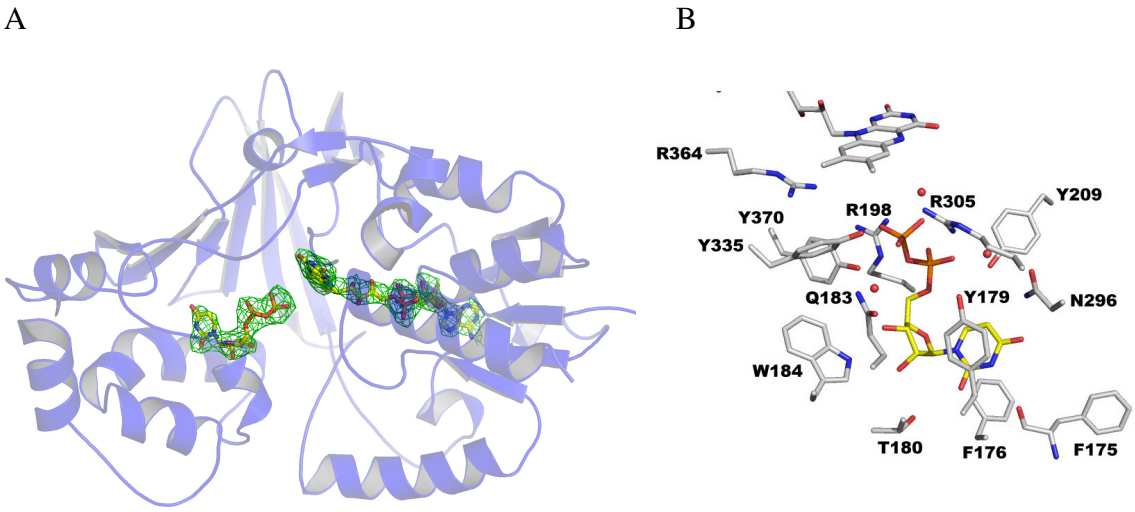


Figure 4-13 Structure of drUGM:UDP complex. **A)** Ribbon diagram representation of the overall structure of drUGM:UDP. The simulated annealing omit maps (contoured to 2.0 σ) for FAD and UDP were shown as green mesh. FAD and UDP were modeled into the density. The two mobile loops exist in a closed conformation. **B)** Binding mode of UDP. Active site interactions of UDP is comparable to the interactions of UDP moiety in drUGM:substrate complexes. The sugar binding region is occupied by a water molecule (red sphere).

The overall structure of drUGM:UDP complex is similar to drUGM$_{ox}$ and drUGM$_{red}$ including the conformation of two flexible loops (0.54 Å and 1.0 Å rmsd for drUGM$_{ox}$ and drUGM$_{red}$ respectively). UDP is entirely buried in the active site and is folded in a conformation similar to the UDP portion of the substrate. The active site interactions of UDP are similar to the UDP moiety of the substrate found in the drUGM$_{ox}$ and drUGM$_{red}$ (Figure 4-13). The sugar binding region is occupied by a water molecule. Binding of UDP in the closed conformation of drUGM indicates that active site closure is

driven by interactions with the UDP portion of the substrate, which is consistent with previous studies.

4.9 Structure of drUGM:UMP complex

The structure of drUGM:UMP complex was determined in order to understand the role of β-phosphate in UDP and its effect on loop movement and the active site interactions.

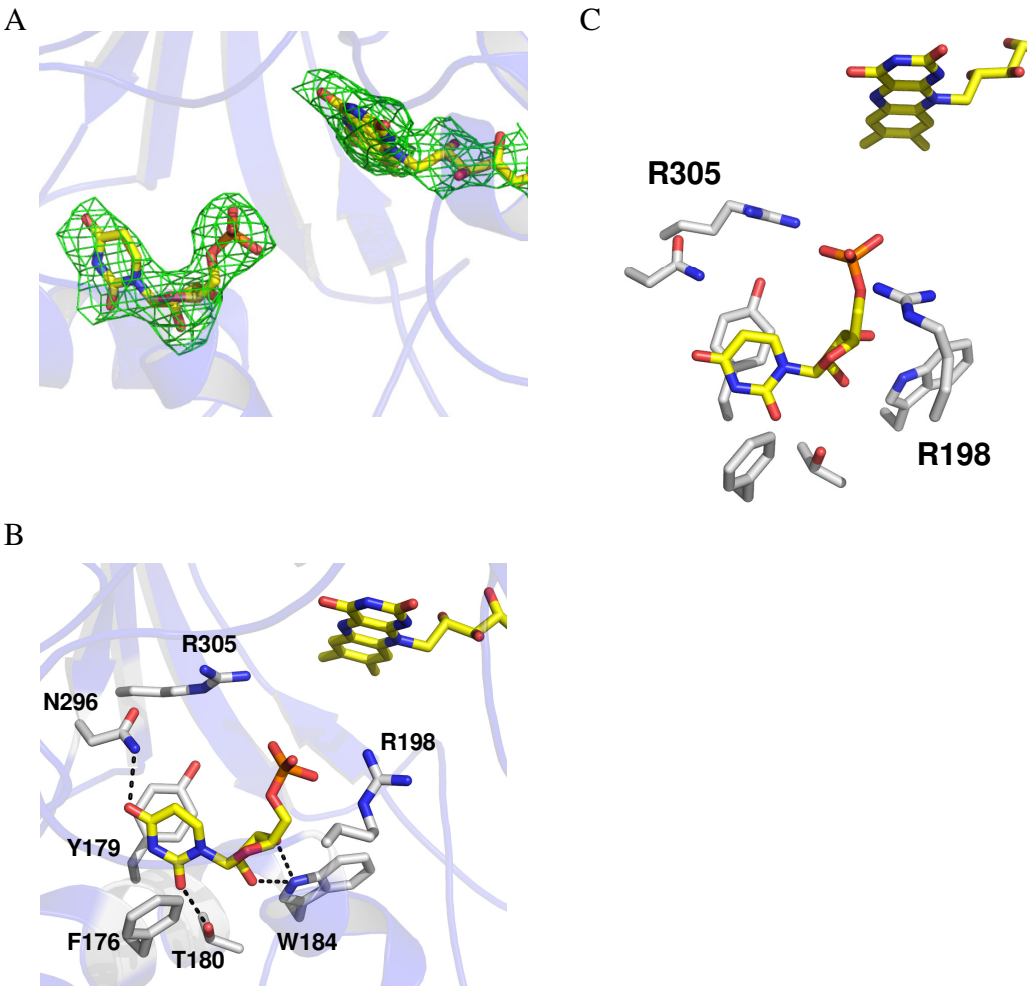


Figure 4-14 Structure of drUGM:UMP complex. **A)** Simulated annealing omit maps (contoured to 2.0 σ) with UMP and FAD modeled into the density. **B)** Active site interactions of UMP is similar to the UMP portion of UDP from drUGM:UDP complex. Arg305 is located away from the phosphate group of UMP. **C)** Interactions of UMP in chain B. Arg305 is oriented close to the phosphate group of UMP.

crystals were grown by the microbatch method and crystallized in the $P2_12_12_1$ space group. Crystals of drUGM:UMP complexes diffracted to 2.8 Å and the unit cell dimensions were similar to the drUGM:UDP complex. The structure of the drUGM:UMP complex was determined by using $drUGM_{ox}$ (after removal of ligands) as the starting model for refinement. Data collection and refinement statistics are shown in Table 4-5. The asu has 5 homodimers and the overall structure is similar to the drUGM:UDP complex structure.

Table 4-5 Data collection statistics and refinement for drUGM:UMP complex.

	drUGM:UMP
Data collection	
Space group	$P2_12_12_1$
Cell dimensions	
a, *b*, *c* (Å)	137.8, 175.9, 222.0 90.00, 90.00, 90.00
Resolution (Å)	2.80 (2.90-2.80)
R_{sym}	19.2 (54.6)
$I / \sigma I$	5.6 (2.5)
Completeness (%)	100 (100)
Redundancy	7.5
Refinement	
Resolution (Å)	30.0-2.8
No. reflections	964991
R_{work} / R_{free}	0.21/0.27
No. atoms	
Protein	29763
Ligand (FAD and UMP)	740
Water (molecules)	88
B-factors	
Protein	58.0
Ligand	67.4 (UMP) 62.3 (FAD)
Water	48.9
r.m.s. deviations	
Bond lengths (Å)	0.007
Bond angles (°)	1.1

*values in the parenthesis are for the highest resolution shell

The simulated annealing difference map revealed electron density for FAD and UMP (Figure 4-14). The two mobile loops exist in the closed conformation. Arg198 of the mobile loop stabilizes the phosphate group of UMP, but Arg305 is located away from the phosphate group (Figure 4-14). Minor differences in the side chain orientations for Arg198 and Arg305 were observed in chain B of drUGM:UMP complex. In chain B, Arg305 is oriented towards the phosphate group of UMP (Figure 4-14). The side chain of Arg198 showed a different orientation but is able to stabilize the phosphate group of UMP. The significance of these differences in arginine side chain orientation (in chain B) is not clear at this moment.

4.10 Structure of drUGM:UDP-Glc complex

Despite the structural similarity between UDP-Gal*p* and UDP-Glc, the latter is neither a substrate nor an inhibitor for UGM and binds very poorly (K_d = 750 μM for reduced kpUGM) (Gruber et al, 2009a). The structure of drUGM:UDP-Glc complex was determined with the aim of understanding its binding mode, active site interactions (C4 hydroxyl group), effect on mobile loop movement and correlate this structural information with the observed activity of UDP-Glc. Crystals of drUGM:UDP-Glc complex were grown by microbatch using conditions similar to drUGM:substrate complexes. The space group and unit cell dimensions of drUGM:UDP-Glc complex were similar to the drUGM:substrate complex crystals. Crystals of drUGM:UDP-Glc complex diffracted to 2.5 Å. The structure of drUGM:UDP-Glc was determined by using $drUGM_{ox}$ (ligands removed) as the starting model for refinement. Data collection and refinement statistics for drUGM:UDP-Glc complex shown in Table 4-6. The asu has 5 homodimers and the overall structure is similar to drUGM:substrate complex structure.

The simulated annealing difference map reveled electron density for FAD and nucleotide portion (UDP) of UDP-Glc. The density for the UDP portion (of UDP-Glc) was well defined in all the monomers. Further rounds of model building and refinement showed improvement in density (Figure 4-15) for the sugar moiety in chain D as compared to other chains (poor density in other chains). Hence, UDP-Glc was modeled only in chain D and further rounds of refinement were carried out. The overall structure of drUGM:UDP-Glc complex is similar to drUGM:substrate complex structures, including

Table 4-6 Data collection and refinement statistics for drUGM:UDP-Glc complex.

	drUGM:UDP-Glc
Data collection	
Space group	$P2_12_12_1$
Cell dimensions	
a, *b*, *c* (Å)	137.8, 175.9, 222.0 90.00, 90.00, 90.00
Resolution (Å)	2.50 (2.59-2.50)
R_{sym}	16.7 (59.9)
$I / \sigma I$	5.2 (2.0)
Completeness (%)	96.6 (80.0)
Redundancy	7.4
Refinement	
Resolution (Å)	20-2.5
No. reflections	189638
R_{work} / R_{free}	0.21/0.27
No. atoms	
Protein	30186
Ligand (FAD, UDP and UDP-Glc)	791
Water (molecules)	525
B-factors	
Protein	54.4
Ligand	62.8 (UDP-Glc) 58.6 (UDP) 58.0 (FAD)
Water	52.3
r.m.s. deviations	
Bond lengths (Å)	0.008
Bond angles (°)	1.2

*values in the parenthesis are for the highest resolution shell

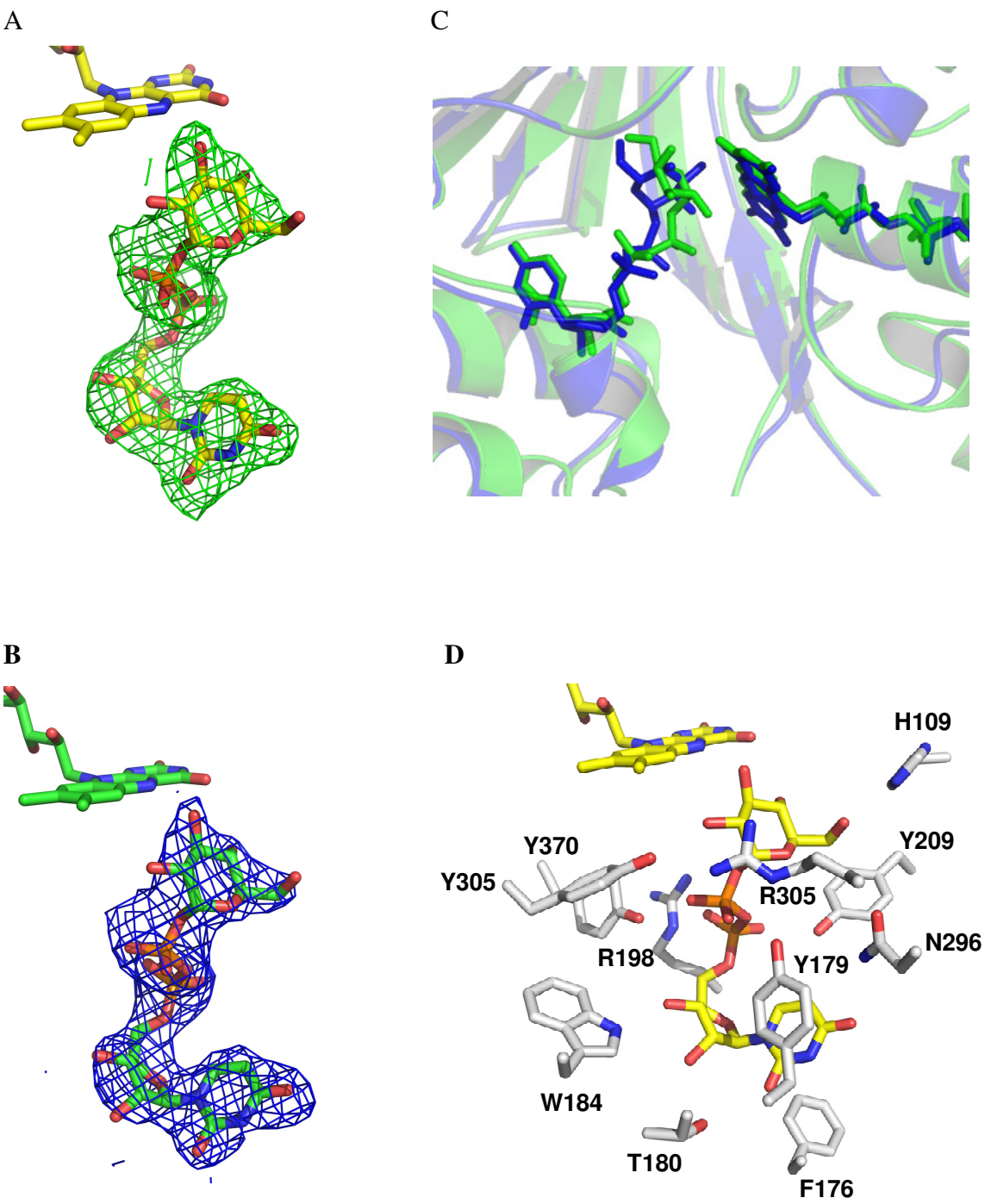


Figure 4-15 Structure of drUGM:UDP-Glc complex. **A)** Simulated annealing omit map (contoured to 2.0 σ) for UDP-Glc in chain D of drUGM:UDP-Glc complex. UDP-Glc modeled into the density. **B)** $2F_o$-F_c map (contoured to 1.0 σ) of UDP-Glc after refinement. **C)** Overlay of $drUGM_{ox}$ (green) and UDP-Glc complex (blue). The overall structure of drUGM:UDP-Glc is similar to $drUGM_{ox}$ including the two mobile loops. Note the differences in the orientation of sugar moiety in $drUGM_{ox}$ and drUGM:UDP-Glc complex. **D)** Active site interactions of UDP-Glc.

the conformation of the two mobile loops. However, the binding mode of UDP-Glc is significantly different when compared to the binding mode of UDP-Gal*p*.

4.11 Comparison of drUGM-Glc and drUGM:UDP-Gal*p* complex structures

A detailed comparison of drUGM:UDP-Glc complex structure with drUGM:UDP-Gal*p* structures revealed key differences in the binding mode and active site interactions (Figure 4-16). The binding mode of UDP-Glc is significantly different from the binding mode of UDP-Gal*p* except in the uridine binding region. Stacking interactions for the uracil base and the hydrogen bonding pattern of the ribose moiety were similar in drUGM:substrate structures and the UDP-Glc complex structure (Figure 4-15). The orientations of the two phosphate groups in UDP-Glc are considerably different than the phosphate groups of UDP-Gal*p* (Figure 4-16). The phosphate groups of the UDP-Gal*p* (in drUGM:UDP-Gal*p* structures) are in close contact with the conserved arginines (Arg198 and Arg305). In the drUGM:UDP-Glc complex structure, Arg198 is in close contact with the α-phosphate group of UDP-Glc, but Arg305 is oriented towards the sugar moiety of UDP-Glc. The side chain of Arg305 is tilted at an angle of ~ 45^{o} as compared to its orientation in the substrate complex structure. This could result in poor or weak stabilization of the β-phosphate of UDP-Glc by Arg305. Furthermore, the anomeric carbon of galactose moiety in $drUGM_{red}$ is located at a distance of 2.80 Å from N5 of reduced FAD (a productive distance) (Figure 4-15) to form FAD-galactosyl adduct. The corresponding distance for glucose in drUGM:UDP-Glc complex is 5.8 Å (a non-productive distance). The orientation of the glucose moiety is significantly different compared to the galactose moiety of the substrate complex structures. In the drUGM:UDP-Glc complex, the C2 hydroxyl group (of glucose) forms a hydrogen bond

with the side chain of Tyr335 and the C3 hydroxyl group is located at a distance of 2.8 Å from N5 of FAD (Figure 4-15). The C4 hydroxyl group is located at a distance of 3.3 Å from O4 of FAD and its equatorial orientation results in weak hydrogen bond interactions

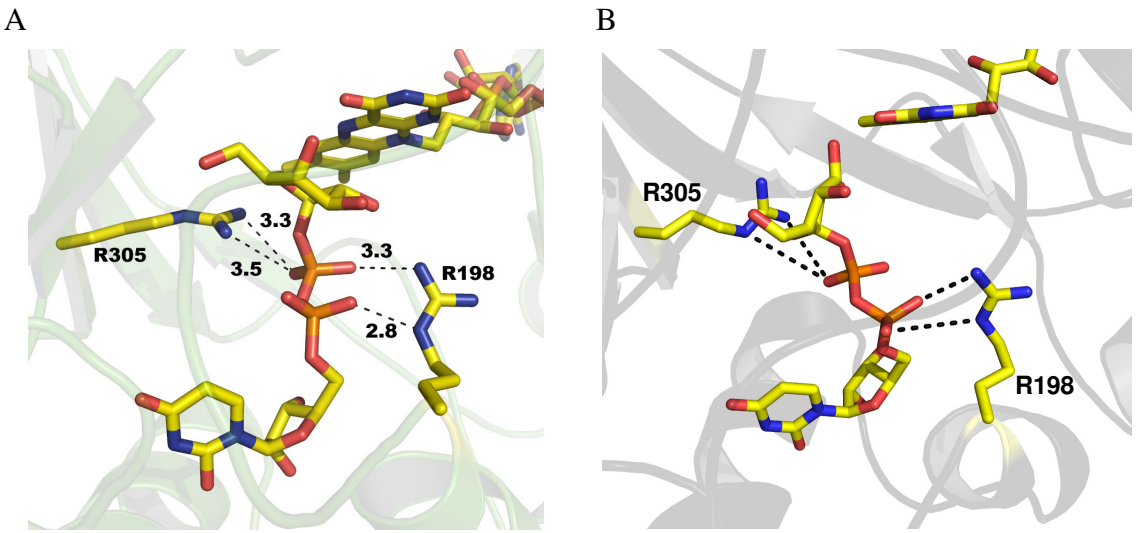


Figure 4-16 Comparison of the diphosphate interactions of drUGM$_{ox}$ and drUGM:UDP-Glc. **A)** In drUGM$_{ox}$ the two phosphate groups of UDP-Gal*p* are stabilized by Arg198 and Arg305. **B)** In drUGM:UDP-Glc complex the side chain Arg305 is tilted and is not in close contact with the β-phosphate of UDP-Glc, however, Arg198 is able to interact with the α-phosphate of UDP-Glc. Also, note the orientation of the glucose moiety as compared to the galactose moiety in A.

with O4 of FAD. In addition, the C4 hydroxyl of glucose orients towards Phe210 and Tyr209, an unfavorable hydrophobic environment. The C6 hydroxyl group of glucose forms weak hydrogen bond interactions with His109 and this interaction was also observed substrate complex structure. The altered binding mode and interaction pattern of UDP-Glc provide a basis for the observed differences in the recognition between UDP-Glc and UDP-Gal*p*. These structural observations provide a basis for the inability of UDP-Glc to act as a substrate or inhibitor for UGM.

4.12 Comparison of drUGM:UDP-Glc with kpUGM:UDP-Glc complex structure

During the progress of our structural investigations on drUGM, the structure of kpUGM in complex with UDP-Glc was reported (Gruber et al, 2009a). Comparison of

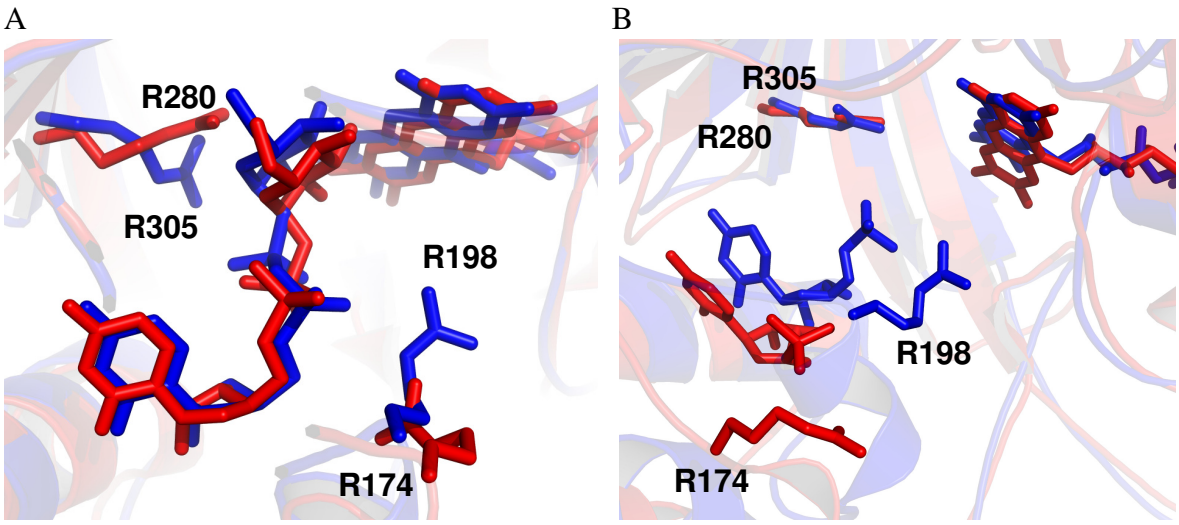


Figure 4-17 Comparison of drUGM:UDP-Glc and kpUGM:UDP-Glc complex. **A)** Overaly of drUGM:UDP-Glc (blue) and kpUGM:UDP-Glc (red) structures. The binding of the UDP-Glc is similar in both the structures. The conserved arginines show differences in their interaction pattern with the phosphate groups of the UDP-Glc. Note the orientations of Arg198 vs Arg174 and Arg305 vs Arg280. **B)** Overlay of the monomer B of kpUGM:UDP-Glc with drUGM:UMP complex. Monomer B of kpUGM:UDP-Glc has only UMP bound to it and hence we compared this with drUGM:UMP complex. The alignment suggests that UMP is bound in a different conformation and location in kpUGM as compared to the binding mode of UMP in drUGM.

drUGM:UDP-Glc and kpUGM:UDP-Glc complex structures revealed significant differences between them. The overall structures of drUGM:UDP-Glc and kpUGM:UDP-Glc are similar, but differences in active site interactions and mobile loop conformations were observed (Figure 4-17). The mobile loops in kpUGM:UDP-Glc exist in a different conformation (partially closed) compared to the fully closed conformation observed in drUGM:UDP-Glc complex structure. As a result, stabilization of α-phosphate of UDP-Glc by Arg174 is weak compared to the close contact maintained by the corresponding

residue (Arg198) in drUGM:UDP-Glc complex. The binding mode of UDP-Glc in kpUGM and in drUGM is similar. The active site Arg280 (in kpUGM) is located away from the β-phosphate of UDP-Glc and orients towards the sugar moiety (close to C6 hydoxyl group of glucose). The corresponding residue in drUGM (Arg305) is also oriented towards the sugar moiety, but not close to C6 hydroxyl group of glucose. The Arg305 side chain is tilted slightly and is able to make contacts with the β-phosphate of UDP-Glc. The mobile loops in drUGM:UDP-Glc complex exist in a closed conformation (in all chains). UDP moiety was modelled in chains that did not show density for sugar moiety. In those chains that were modeled with UDP, the interaction pattern is similar to UDP bound structure (drUGM:UDP complex). In kpUGM:UDP-Glc complex, the active site of monomer A and monomer B contains UDP-Glc and UMP respectively. The mobile loop of monomer B exists in an open conformation and the interaction of UMP is significantly different from the UMP portion of UDP-Glc bound to monomer B. The binding mode of UMP (monomer B) is also significantly different from the binding mode of UMP observed in our drUGM:UMP complex structure. For example, the phosphate moiety of UMP is not in close contact with the conserved arginines (Arg174 and Arg284 in kpUGM) compared to the close contacts maintained by Arg198 and Arg305 in drUGM:UMP complex. The differences in the mobile loop conformation between drUGM:UDP-Glc and kpUGM:UDP-Glc complexes may be explained on the basis of the method of crystallization used to form the complex (Gruber et al, 2009a). The kpUGM:UDP-Glc complex crystals were obtained by soaking. UDP-Glc was soaked into preformed unliganded kpUGM crystals. drUGM:UDP-Glc complex was crystallized by cocrystallization method. It is well known that conformational changes may not be

apparent in soaking methods (Hassell et al, 2007). In such cases there is a need to cross validate the structures by forming the protein-ligand complexes by cocrystallization methods. This is evident from the structures of kpUGM:UDP-Glc and drUGM:UDP-Glc obtained by different crystallization methods.

CHAPTER 5

Evaluation of UGM inhibitors

5.0 Ligand-based approach: Phosphonate analog of UDP-Gal*p* (GCP)

Inhibitors of enzymes that mimic the structure of natural substrate are called substrate analogs. They will not undergo any turnover upon binding to the enzyme and inhibit the reaction (except suicide inhibitors). Substrate analogs are examples of ligand-based design. GCP is a phosphonate analogue of UDP-Gal*p* with the anomeric (Figure 5-1) oxygen replaced by a methylene functionality. The rationale behind the design of GCP is based on the previous mechanistic investigations reported for UGM (Barlow et al, 1999; Fullerton et al, 2003; Huang et al, 2003; Soltero-Higgin et al, 2004). It has been proposed that the C-O anomeric bond of UDP-Gal*p* is cleaved during the reaction with the formation of an FAD-galactosyl adduct and UDP (a good leaving group) (See section 1.3, Figure 1-5 and 1-6). The substitution of a carbon atom for the anomeric oxygen (as in GCP) would form a C-glycosidic linkage that cannot be cleaved during the reaction and therefore GCP would be expected to act as an inhibitor of UGM.

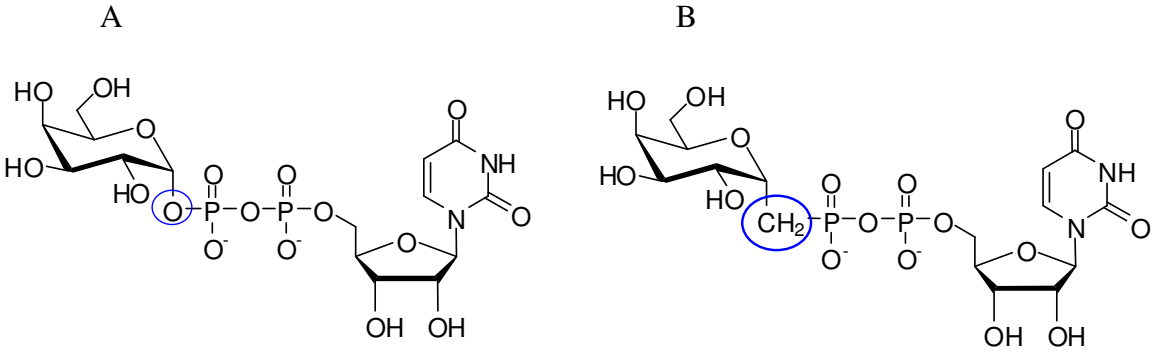


Figure 5-1 Design of phosphonate analog as inhibitor of UGM. **A)** Structure of UDP-Gal*p* **1** to illustrate the design of its phosphonate analog. The anomeric oxygen is highlighted by blue circle. **B)** The phosphonate analog of UDP-Gal*p* (GCP) is designed by replacing the anomeric oxygen by the methylene group (blue circle) at the anomeric position. The phosphonate analog is expected not to be cleaved but inhibit UGM.

Replacing an oxygen atom with a methylene moiety is considered as an isosteric replacement, and has been used widely in medicinal and biological chemistry to probe phosphoryl transfer processes (Engel, 1977; Romanenko & Kukhar, 2006). X-ray analyses of phosphonates show that the bond angles and bond lengths are within 10 % of the natural substrates (Engel, 1977).

Previously, synthesis and activity measurement of GCP and UDP-1C-Gal*f* (phosphonate analog of UDP-Gal*f*) towards *E. coli* UGM (ecUGM) has been reported (Caravano & Vincent, 2009). The inhibitory activity of GCP and UDP-1C-Gal*f* were determined for reduced and oxidized ecUGM. GCP and UDP-1C-Gal*f* showed <10 % inhibition under reducing conditions. GCP and UDP-1C-Gal*f* displayed 38 % and 91 % inhibition under non-reducing conditions respectively. The concentration of substrate (UDP-Gal*f*) used in the evaluation of GCP and UDP-C-Gal*f* was set to 1 mM which is much higher than the K_m of UDP-Gal*f* (27 μM) for ecUGM (Caravano & Vincent, 2009; Zhang & Liu, 2000). This would make proper evaluation of weak inhibitors difficult.

5.1 Inhibitory activity of GCP

To gain insights on the inhibitory activity and binding, GCP was evaluated against drUGM, kpUGM and mtUGM. GCP showed greater than 50% inhibition against all three UGMs, when used at 1 mM (all inhibition studies were performed under reducing conditions). The inhibitory activity of GCP was further evaluated at different concentrations to determine the IC_{50} values against drUGM, kpUGM and mtUGM (Table 5-1). The IC_{50} values of GCP against various UGMs suggest that GCP is a weak inhibitor of UGM.

Table 5-1 Inhibitory activity of GCP towards kpUGM, mtUGM and drUGM.

Enzyme	IC_{50} (μM)*
kpUGM	**479 ± 34**
mtUGM	**495 ± 38**
drUGM	**411 ± 13**

*The errors represent the standard deviation based on duplicate measurements

5.2 Structure of drUGM:GCP complex

In order to determine the binding mode and active site interactions of GCP, the UGM:GCP complex structure was determined. The substrate binding mode in drUGM is similar to the kpUGM:UDP-Gal*p* reduced complex structure (Gruber et al, 2009b) and the inhibition profile of GCP towards drUGM, kpUGM and mtUGM are similar. Also, I had high success rate in forming ligand complexes using drUGM. Therefore, I decided to use drUGM to understand GCP binding and its active site interactions. The structure of the drUGM:GCP complex is the first report of a complex structure of a phosphonate analog of UDP-Gal*p* bound to an enzyme (PDB code 3MJ4). The drUGM:GCP complex was crystallized by the microbatch method and the crystals diffracted to 2.6 Å. The unit cell dimensions and space group of drUGM:GCP complex were similar to drUGM:substrate complex crystals. Data collection and refinement statistics are shown in Table 5-2. The drUGM:GCP structure determined by using $drUGM_{ox}$ (without ligands) as the starting model for refinement. The simulated annealing difference map revealed well defined density for the ligands (FAD and GCP). But, the density for GCP is complete only in chain D (Figure 5-2). Hence, GCP was modeled only in chain D. In other chains the density for sugar moiety was weak and we modeled UDP in those chains. The overall

structure of the drUGM:GCP complex (including the two mobile loops) is similar to the drUGM:UDP-Gal*p* complex structures (Figure 5-3). The rmsd between the drUGM:UDP-Gal*p* complex and the drUGM:GCP complex is 0.4 Å for all equivalent Cα atoms. Based on the structural similarity (between UDP-Gal*p* and GCP), it is anticipated that the binding mode of GCP would be similar to UDP-Gal*p*. However,

Table 5-2 Data collection and refinement statistics for drUGM:GCP complex.

Data collection	**drUGM:GCP complex**
Space group	$P2_12_12_1$
Cell dimensions	134.1, 175.6, 223.9
a, *b*, *c* (Å)	90.00, 90.00, 90.00
Resolution (Å)	2.65 (2.74-2.65)
R_{sym}	14.9 (57.8)
$I / \sigma I$	6.6 (2.0)
Completeness (%)	100 (100)
Redundancy	7.1 (7.0)
Refinement	
Resolution (Å)	44.5-2.65
No. reflections	162540
R_{work} / R_{free}	0.20/0.26
No. atoms	
Protein	29702
Ligands	755
Water (molecules)	613
B-factors	
Protein	53.3
Ligand	62.6
Water	45.3
r.m.s. deviations	
Bond lengths (Å)	0.008
Bond angles (°)	1.2

*values in the parenthesis are for the highest resolution shell

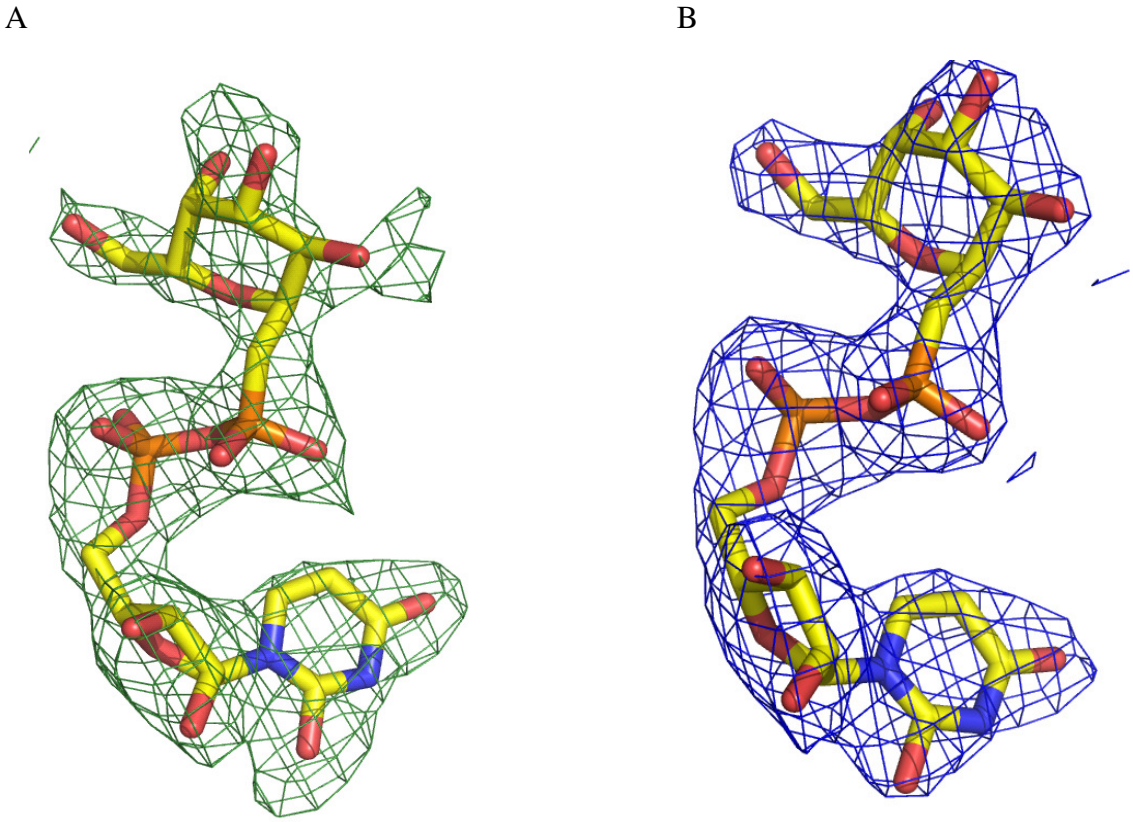


Figure 5-2 Electron density maps of GCP in drUGM:GCP complex structures. **A)** The omit map of GCP in chain D (contoured to 2.0 σ) after simulated annealing refinement. GCP molded into the density after simulated annealing refinement. **B)** The $2F_o$-F_c map of GCP (contoured to 1.0 σ) in chain D after refinement.

differences in the binding mode and active site interactions were observed for GCP. GCP is bound in a different folded conformation (Figure 5-3) as compared to the U-shaped conformation of UDP-Gal*p* observed in the substrate complex structures (Partha et al, 2009). The interactions of GCP at the uridine binding region (Figure 5-3) were similar to the uridine moiety of UDP-Gal*p* in drUGM:UDP-Gal*p* structures. Notable differences were observed in the diphosphate and sugar binding regions. The β-phosphate group of GCP is bent on top of the uridine, resulting in significant alterations in the orientation of

the sugar moiety and the interaction patterns of the β-phosphate group (Figure 5-3). In the overlaid structures of GCP (from drUGM:GCP complex) and UDP-Gal*p* (from drUGM:UDP-Gal*p* complex) the distance between the phosphorous atom of α-phosphate and β-phosphate group in GCP to the corresponding atoms of UDP-Gal*p*

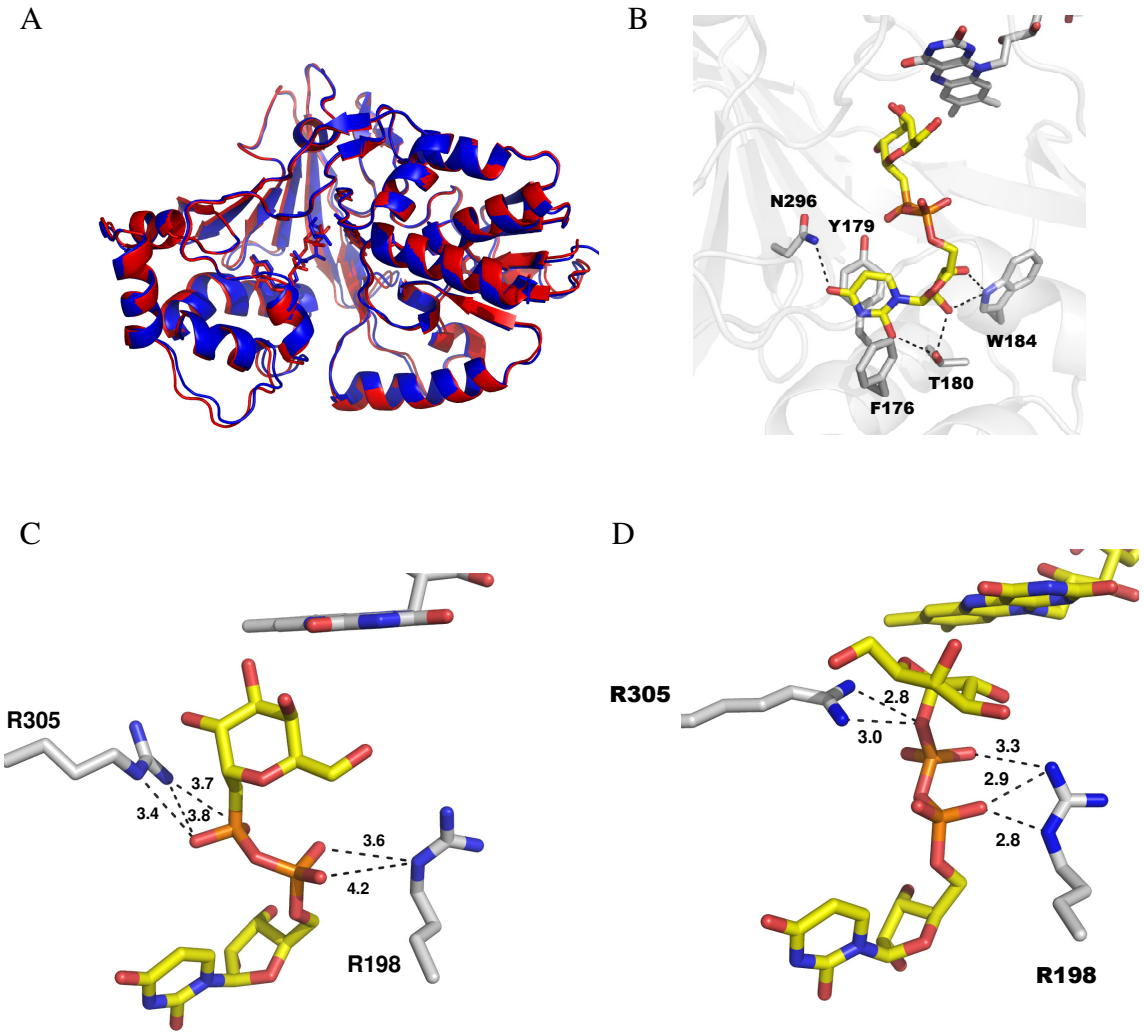


Figure 5-3 Overall structure of drUGM:GCP complex and the active site interactions. **A)** Overlay of drUGM:GCP (Red) and drUGM:UDP-Gal*p* (Blue) structures. **B)** Interactions of the uridine moiety of GCP in drUGM:GCP structure. **C)** Orientation of the α-phosphate and β-phosphono groups of GCP and their interactions with R305 and R198. **D)** Interactions of the two phosphate groups of UDP-Gal*p* with R305 and R198. Note the distance between the anomeric oxygen and NH1/NH2 of Arg305.

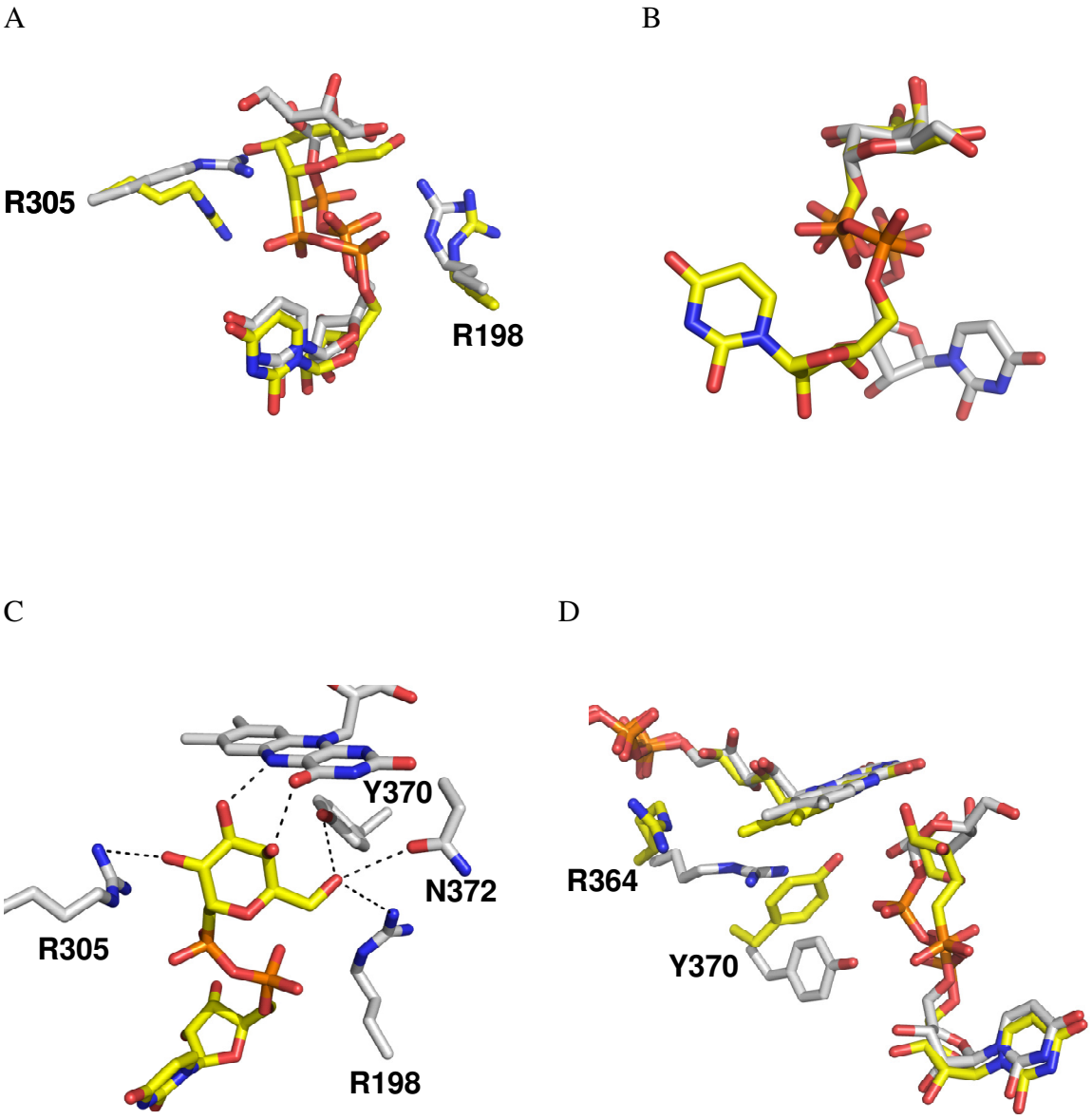


Figure 5-4 Orientation of the conserved arginines and the sugar moiety in drUGM:GCP complex. **A)** Overlay of drUGM:GCP (Carbon, gray) and drUGM:UDP-Gal*p* (Carbon, yellow) structures to highlight the orientation of the two active site arginines (R305 and R198). The alignment also reveals the differences in the sugar orientation between the two complex structures. **B)** Alignment of bound conformation of GCP (Carbon, yellow) from drUGM:GCP complex and UDP-Gal*p* (Carbon, gray) from drUGM:UDP-Gal*p* complex to emphasize that the sugar moiety is able to maintain the $^4C_1$ conformation in both the complexes. **C)** Interactions of the sugar moiety of GCP in drUGM:GCP complex. **D)** Overlay of drUGM:GCP (Carbon, yellow) and drUGM: UDP-Gal*p* (Carbon, gray) structures to highlight the differences in the side chain orientation of Arg364 and Tyr370.

was found to be 1.1 Å and 2.5 Å respectively. In addition, the distance between the phosphorus atom of β-phosphate (in UDP-Gal*p*) and the uracil ring was 7.1 Å in the drUGM:UDP-Gal*p* complex. The corresponding distance for GCP in the drUGM:GCP complex was 5.0 Å. These measurements clearly indicate that the β-phosphate group of GCP undergoes significant conformational changes upon binding to UGM. Structural and biochemical studies revealed the role of two conserved active site arginines that stabilize the two phosphate groups of the UDP-Gal*p*. A closer look at the interactions of β-phosphate in drUGM:UDP-Gal*p* complex structure revealed that the anomeric oxygen atom is located at a distance of 2.8 Å and 3.0 Å from NH1 and NH2 of Arg305 respectively (Figure 5-3). In GCP, the anomeric oxygen is substituted with a methylene group that could extend the bond lengths and alter the bond angles (by up to 10 %) and would reduce the electronegativity at that position. In the drUGM:GCP structure, the side chain of Arg305 adopts a folded conformation as compared to the extended orientation of Arg305 side chain observed in the drUGM:UDP-Gal*p* structure. The CD-NE-NE-CZ fragment of Arg305 in drUGM:GCP is tilted ~ 45^{o} to the corresponding fragment in drUGM:UDP-Gal*p* complex. This orientation would avoid any steric clashes with the methylene group of GCP (Figure 5-4). The rearrangement of Arg305 side chain and the bent conformation of the β-phosphate group of GCP could be due to unfavorable steric interactions between the methylene group and the side chain of Arg305. The α-phosphate group of GCP showed only subtle changes in its geometry and is located at a stabilizing distance from Arg198 (Figure 5-3). The NH1 and NH2 of Arg198 is tilted at a different angle when compared with the drUGM:substrate complex structure. In drUGM:UDP-Gal*p* complex structure, Arg198 stabilizes the α-phosphate group of UDP-Gal*p* and is

also in close contact with the one of the oxygen atoms of β-phosphate group (Figure 5-3). In the drUGM:GCP structure the bent geometry of the β-phosphate group result in the loss of additional stabilization by Arg198.

It is possible that any distortion of the β-phosphate group would alter the binding mode of the hexose ring. As expected, the galactose moiety of GCP is oriented in a different fashion and the interactions (Figure 5-4) are different compared to the galactose portion of the substrate in the drUGM:substrate structure (Partha et al, 2009). The sugar moiety is able to maintain the $^4C_1$ conformation (Figure 5-4). In the drUGM:GCP structure, the C2 hydroxyl group (of galactose) is oriented close to the side chain of Arg305 and the C3 hydroxyl group forms a hydrogen bond with the N5 of FAD. The corresponding hydroxyl groups of galactose in the drUGM:UDP-Gal*p* complex form water mediated hydrogen bonds with His88, Asn372, Arg364 and Tyr371 (Partha et al, 2009). The C6 hydroxyl group of GCP forms hydrogen bonds with the side chains of Asn372, Tyr370 and Arg198. In UDP-Gal*p* (from the drUGM:UDP-Gal*p* structure) the C6 hydroxyl group forms a hydrogen bond with His109. The C4 hydroxyl group of GCP forms a hydrogen bond with the O4 of FAD and this interaction is also observed in the drUGM:UDP-Gal*p* complex structures.

The drUGM:GCP complex structure reveals other conformational changes when compared to the drUGM:UDP-Gal*p* structures. The side chains of Arg364 and Tyr370 are oriented towards the substrate binding pocket in the drUGM:UDP-Gal*p* structure, whereas, in the drUGM:GCP structure the side chains of Arg364 and Try371 are in close contact with the phosphate group of FAD and N10 of FAD respectively (Figure 5-4). The binding mode and interactions of GCP in the drUGM:GCP complex structure suggest that

subtle isosteric changes in the ligand structure can result in significant alterations in its binding mode with UGM.

5.3 Comparison of the binding mode of GCP with other sugar nucleotide phosphonates

There are few other reported structures of enzymes complexed with sugar nucleotide phosphonates (Clarke et al, 2008; Gordon et al, 2006). The crystal structures of UDP-N-acetylglucosamine (UDP-GlcNAc) and the corresponding phosphonate analog, (UDP-1C-GlcNAc) in complex with rabbit N-acetylglucosaminyltransferase 1 (GnT1) show no differences in the binding mode of the phosphonate analog compared to the natural substrate (PDB code, 2APC). The K_i of the phosphonate analog towards GnT1 was found to be 28 μM, consistent with a similar binding mode to the natural substrate (K_m = 78 μM). Analysis of the active site interactions of GnT1 with the natural substrate revealed that the hydroxyl of Ser322 forms a hydrogen bond with the anomeric oxygen, whereas in the phosphonate complex, this interaction is missing and the hydroxyl group rotates away to form a hydrogen bond with backbone NH of Ser322. The active site architecture of GnT1 does not discriminate between the isosteric changes to ligand structure (oxygen *vs.* methylene), resulting in identical binding modes for substrate and the phosphonate analog.

The other sugar nucleotide phosphonate structure is the complex between the C-1 phosphonate analog of UDP-GlcNAc and O-linked GlcNAc transferase (OGT) which is involved in the post-translational modification of proteins (PDB code 2JLB). This phosphonate analog is a poor inhibitor of OGT, IC_{50} > 5 mM (K_m for natural substrate is 0.5 μM) (Hajduch et al, 2008). The poor inhibitory activity of this compound was

attributed to the rigid active site architecture of OGT that is able to discriminate between the change in geometry of *O*-glycosidic bond and the *C*-glycosidic bond. However,

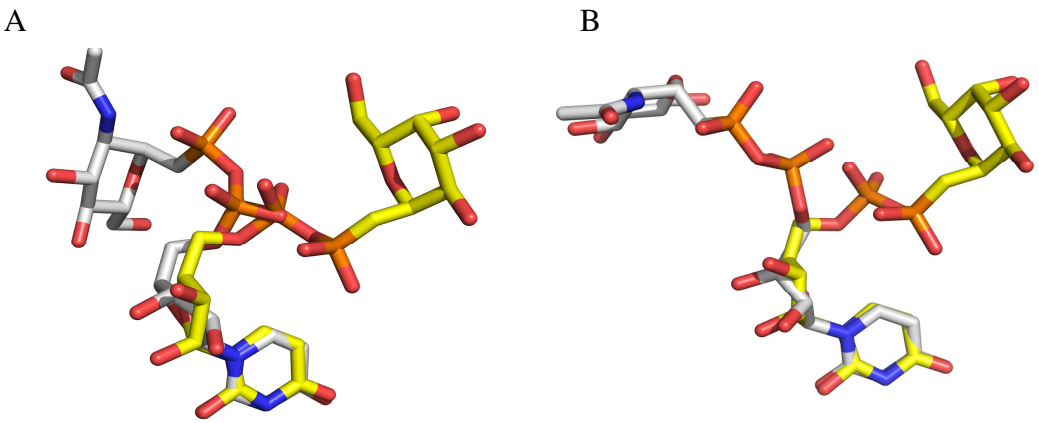


Figure 5-5 Overlay of binding conformation GCP (from drUGM:GCP complex) with UDP-1C-GlcNAc. **A)** Alignment of GCP (Carbon, yellow) and UDP-1C-GlcNAc (Carbon, gray) from GnT1 (PDB code 2APC). **B)** Alignment of GCP (Carbon, yellow) and UDP-1C-GlcNAc (Carbon, gray) from OGT (PDB code 2JLB). The urdine portion of GCP and UDP-1C-GlcNAC aligns well (A and B) compared to the diphosphate and sugar region of the molecule. Note the proximity of β-phosphate group of GCP to the uracil ring, such an orientation not observed for UDP-1C-GlcNAc in A and B.

structural and site-directed mutagenesis studies suggest that the phosphonate analog binds in a similar mode to the natural substrate (Clarke et al, 2008), although the crystal structure of OGT complexed with UDP-GlcNAc has not been published.

Alignment of GCP with the phosphonates from 2APC and 2JLB was performed (Figure 5-5) to examine the differences in the binding mode of GCP and the phosphonate analog of UDP-GlcNAc. The uridine portion of GCP and UDP-1C-GlcNAc (in 2APC and 2JLB) aligned well, but the diphosphate and hexose portions of the molecule adopt vastly different orientations. This comparison reveals that GCP adopt a novel binding conformation not observed in the X-ray crystallographic analysis of the phosphono analogues of UDP-GlcNAc complexed with either GnT1 or OGT.

The weak IC_{50} values of GCP can be explained on the basis of the observed structural reorganization in the active site that occurs upon binding of GCP and an associated loss of favorable interactions, including the stabilization of the two phosphate groups of GCP by the conserved arginines. A similar binding mode and interaction pattern may account for the poor inhibitory activity of GCP and UDP-C-Gal*f* towards ecUGM (Caravano & Vincent, 2009). The drUGM:GCP complex structures provided a structural basis to understand the interactions of GCP at a molecular level and underline the consequence of replacing the anomeric oxygen of the substrate with a methylene functionality. Although, phosphonates are well established isosteric replacements for phosphate groups, the results from this study reveal that in some cases (for example, UGM) such isosteric modifications may not be applicable for inhibitor development. The structural results also illustrate one potential problem in using substrate analogues as probes for the binding of substrates to enzymes in some cases.

5.4 Structure-based approach: Inhibitors from virtual screening

In addition to ligand-based approaches, inhibitors for enzymes can be identified by structure-based methods, such as structure-based virtual screening (SBVS). If the 3-dimensional structure of the enzyme is known, then screening of chemical libraries against the enzyme can be performed by SBVS (De Azevedo, 2010; Ghosh et al, 2006; Lyne, 2002; Mestres, 2002; Sousa et al, 2010; Xu, 2002). On the other hand, if the structure of the target protein is not available or is unknown, then a model of the protein can be prepared by homology modeling for SBVS.

SBVS involves docking (binding mode prediction) and scoring (ranking affinity) of a molecule (from a chemical database) into the binding site of the target protein

(Bissantz et al, 2000). SBVS is complementary to high-throughput screening (HTS) technique (Mestres, 2002) and has been proven to be successful in identifying inhibitors for various therapeutic targets (Lyne, 2002; Sousa et al, 2010; Villoutreix et al, 2009) (Table 5-3). The flow chart for a typical SBVS is shown in Figure 5-6. Various docking

Table 5-3 Examples for successful application of structure-based virtual screening.

Target	Method	Reference
Carbonic anhydrase II	FlexX	(Grüneberg et al, 2002)
Ptp1b	DOCK	(Doman et al, 2002)
Estrogen receptor	PRO_LEADS	(Baxter et al, 2000)
Thrombin	PRO_LEADS	(Baxter et al, 2000)
Factor Xa	PRO_LEADS	(Baxter et al, 2000)
Thymidylate synthase	DOCK	(Tondi et al, 1999)
Retinoic acid receptor	ICM	(Schapira et al, 2000)
Farnesyl transferase	EUDOC	(Perola et al, 2000)
Kinesin	DOCK	(Hopkins et al, 2000)
Hypoxanthine-guanine-xanthine phosphoribosyl tansferase	DOCK	(Aronov et al, 2000)
DNA gyrase	LUDI	(Boehm et al, 2000)
HIV-1 RNA transactivation response element	ICM	(Filikov et al, 2000)

programs have been used for SBVS and they differ in their docking algorithm (binding mode prediction) and the scoring functions.

A number of comparative studies have been reported to evaluate the performance of docking programs for virtual screening (Bissantz et al, 2000; Cummings et al, 2005; Kellenberger et al, 2004; Schneider & Böhm, 2002; Schulz-Gasch & Stahl, 2003; Seifert, 2009; Smith et al, 2003; Stahl & Rarey, 2001; Warren et al, 2006; Wilton et al, 2003). FlexX (Kramer et al, 1999; Rarey et al, 1996; Stahl & Rarey, 2001) is one of the docking programs widely used in the SBVS and proven to be successful in identifying inhibitors for various drug targets (Cho et al, 2008; Grüneberg et al, 2002; Kiss et al, 2008; Miguet et al, 2009; Ramsden et al, 2009)

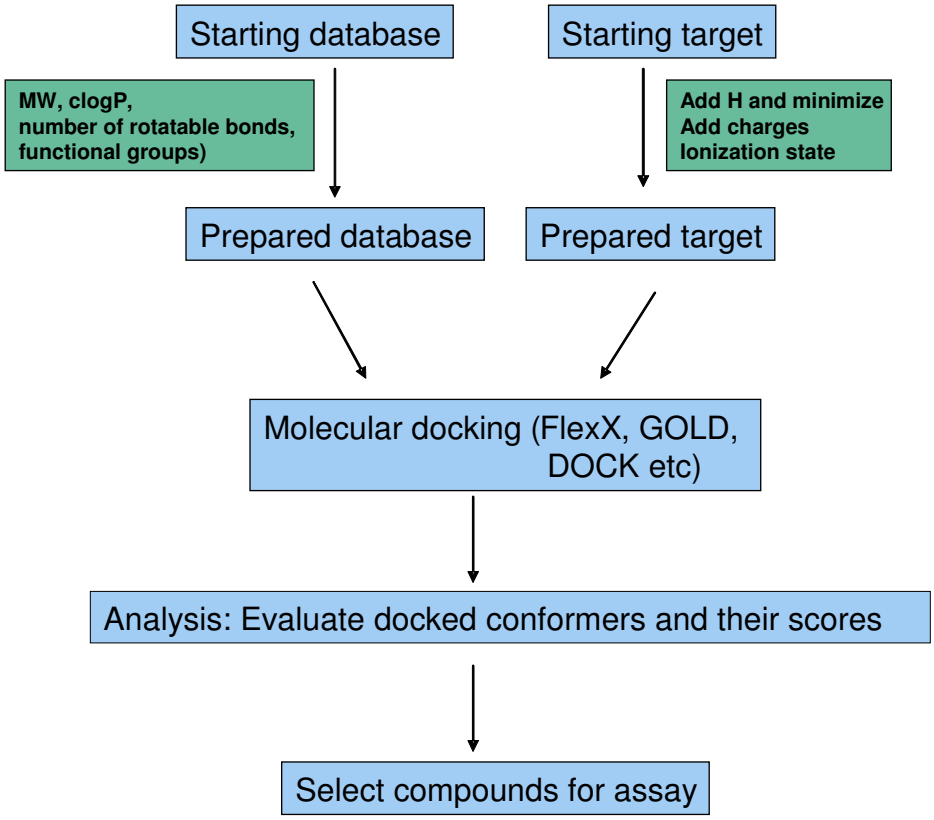


Figure 5-6 Flow chart for structure-based virtual screening of a chemical database.

5.5 Application of SBVS to UGM

Our collaborators (Professor Neil Thomas, Nottingham University) used FlexX (Rarey et al, 1996) docking program to perform SBVS. The crystal structure of kpUGM (Beis et al, 2005) was chosen as the target for SBVS. A structure of the UGM-substrate complex was not known at the time of performing these studies. The LeadQuest database (Tripos Inc.) was used as the virtual chemical library for identifying inhibitors. Based on the results from SBVS, 13 molecules (hits) were purchased and evaluated against various bacterial UGMs (Figure 5-7). The inhibition studies allowed me to identify some novel inhibitors that showed micromolar inhibition. These molecules can be considered as a starting point or lead molecule for further optimization to enhance their potency. The

inhibition studies on UGM demonstrate the utility of SBVS as an alternative to HTS to identify novel enzyme inhibitors.

5.6 Activity of inhibitors from virtual screening

Hits that were identified through SBVS showed promising inhibitory activity towards kpUGM and mtUGM. Initially, inhibition studies were conducted by using 1 mM of inhibitor in assays against kpUGM (Table 5-4). These results were later repeated against mtUGM. Compounds that showed > 80 % inhibitory activity at 1 mM were further screened at lower concentrations of the inhibitor. Due to limited availability of LQ6 and LQ10, they were not screened at 1 mM towards mtUGM. LQ6 and LQ10 showed approximately 50 % inhibition at 1.0 μM (Table 5-5). LQ1 showed 84 % inhibition at 1 mM, but showed only 2 % inhibition at 10 μM against kpUGM. I am unable to screen LQ1 towards mtUGM at lower concentration (due to availability problems). The 13 compounds (Figure 5-7) that were chosen for evaluation can be classified into various structural classes, such as aminosugars, indole analogs, substituted amides, pyrimidine derivative, uridine analog, sugar phosphates, substituted sulfonamides and bridged-sugar analogs. The inhibition profile and the structure-activity relationship of each class of inhibitors are discussed in detail.

5.6.1. Indole analogs

As discussed previously (See section 4.3) the active site of UGM is divided into three regions, each with distinct binding regions, namely, uridine binding region, diphosphate binding region and the sugar binding region (Partha et al, 2009). LQ1, LQ6 and LQ10 showed promising inhibitory activity towards kpUGM and mtUGM. Although LQ1 showed weak inhibition at lower concentrations, it has structural similarity to LQ6

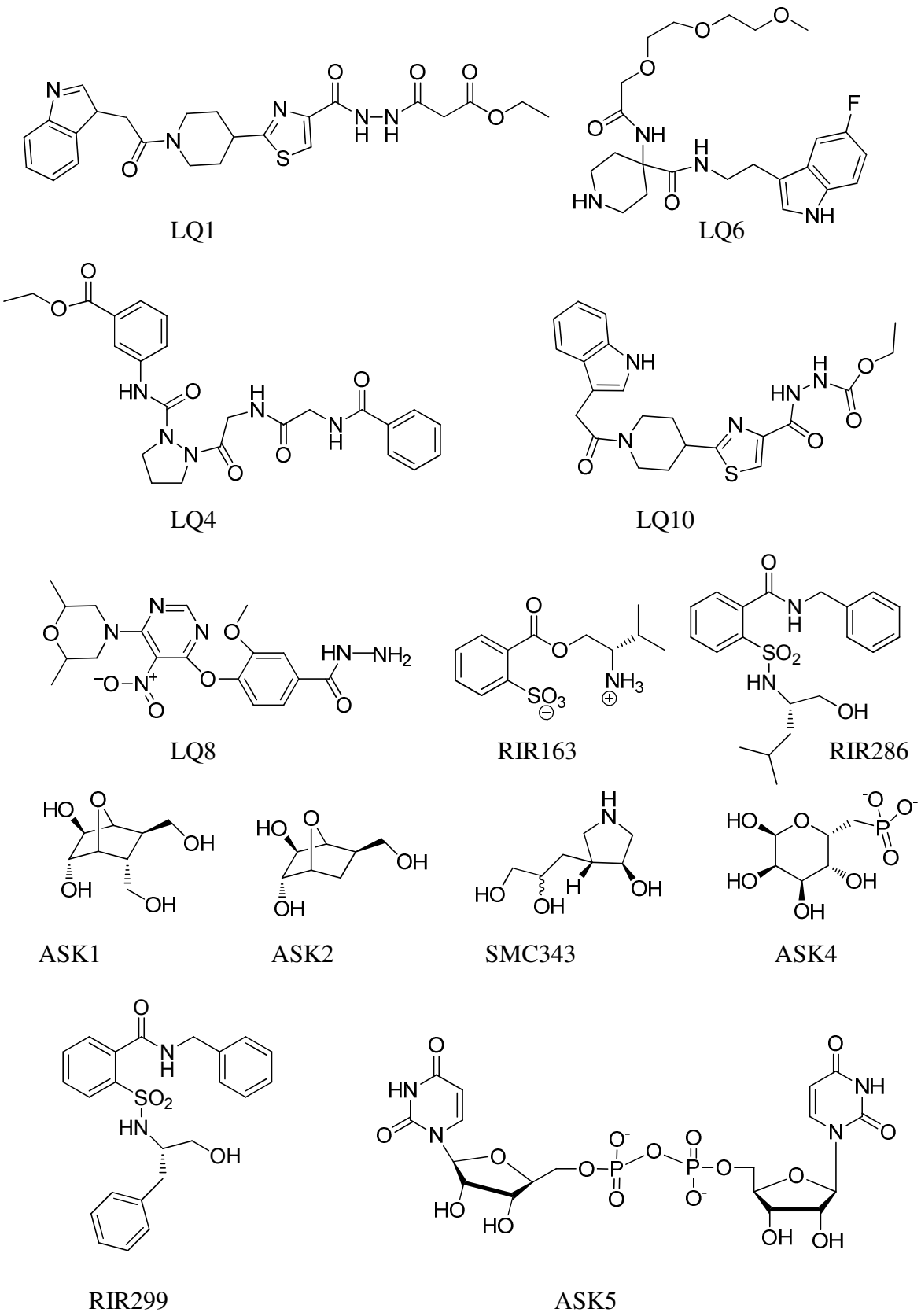


Figure 5-7 structures of the UGM inhibitors identified through virtual screening.

Table 5-4 Inhibition of UGM using 1 mM inhibitors.

Compound	% inhibition	
	kpUGM	mtUGM
LQ1	84 (± 6)*	79 (± 1)
LQ6	82 (± 1)	--
LQ4	77 (± 1)	44 (± 3)
LQ8	84 (± 2	53 (± 1)
LQ10	86 (± 1)	--
SMC343	74 (± 2)	35 (± 2)
RIR163	33 (± 2)	33 (± 2)
RIR299	79 (± 1)	73 (± 4)
RIR286	65 (± 1)	36 (± 4)
ASK1	24 (± 3)	46 (± 3)
ASK2	32 (± 3)	29 (± 2)
ASK4	23 (± 3)	43 (± 4)
ASK5	26 (± 5)	69 (± 2)

-- not determined
* The errors represent the standard deviation based on duplicate measurements

Table 5-5 Inhibition of LQ6 and LQ10 towards kpUGM and mtUGM at 1 μM.

	% inhibition	
	kpUGM	mtUGM
LQ6	52.0 (± 5) *	46 (± 2)
LQ10	48 (± 1)	49 (± 2)

* The errors represent the standard deviation based on duplicate measurements

and LQ10. All three compounds have an indole nucleus and we expected that the indole moiety may form stacking interactions with aromatic residues in the uridine binding region. Docking studies were carried out to predict the binding mode of these inhibitors. SURFLEX-DOCK (Tripos Inc) was used for docking. Coordinates of drUGM:substrate complex used for the docking (UDP-Gal*p* removed from the active). The selection of the docked conformer is based on the stacking interactions with the aromatic residues (Tyr179 and Phe175) and also ability to make contacts with the conserved arginines (Arg198 and Arg305). In the docked conformation of LQ1 and LQ10 the indole moiety is stacked between two aromatic residues (Tyr179 and Phe175) in the uridine binding region (Figure 5-8). The hydrazine moiety and the carbonyl group next to the hydrazine (in LQ1 and LQ10) are in close contact with the side chain of Arg198 (diphosphate binding region). The propyl chain does not show any specific interactions. The carbonyl group that connects the indole ring and the piperidine moiety forms a hydrogen bond with the side chain Trp184 (indole NH). The NH of indole moiety in LQ1 and LQ10 also form hydrogen bonds with side chain of Thr180. LQ6, a closely related molecule to LQ1 and LQ10 showed two different binding modes that are significantly different from LQ1 and LQ10 (Figure 5-8). The two binding modes (BM1 and BM2) did not show any stacking interactions between the indole moiety of LQ6 and Tyr179/ Phe175. In BM1, the indole moiety is stacked in between the isoalloxazine ring of FAD and the side chain of Arg198. The piperidine ring occupies the uridine binding region. The ether linkage attached to the piperidine ring folds on top of the indole moiety. In BM2, the piperidine ring occupies the sugar binding region, and the indole moiety is positioned at the entrance of the active site. The ether chain oriented towards the uridine binding region. The differences in the

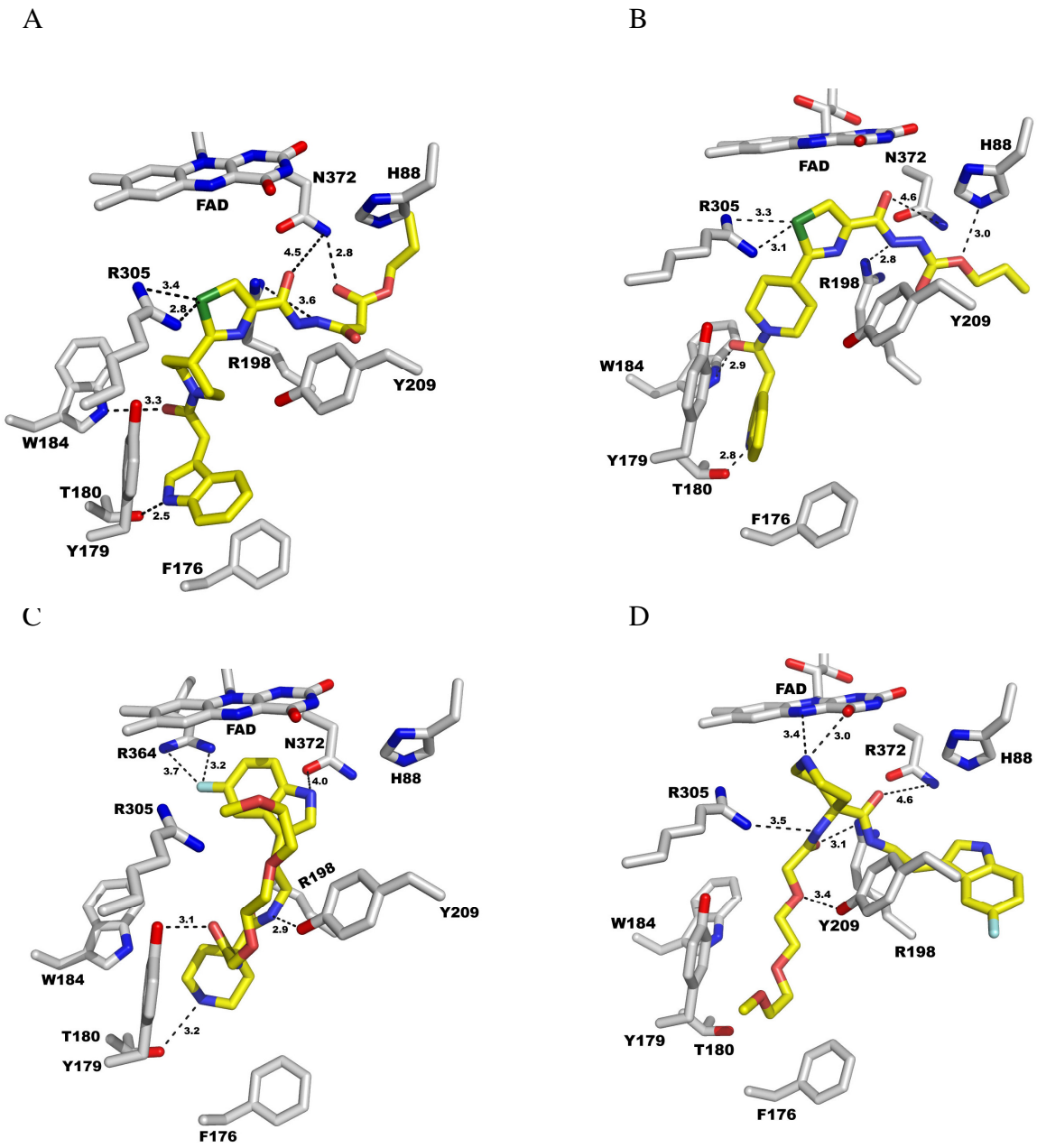


Figure 5-8 Binding mode of indole analogs. Active site interactions of **A)** LQ1, **B)** LQ10, **C)** LQ6- binding mode 1 and **D)** LQ6-binding mode 2.

binding mode of LQ6 compared to LQ1 and LQ10 are likely due to the spiro-type linkage at C4 of piperidine ring and increase in the length of ether chain.

5.6.2. Substituted amide (LQ4) and pyrimidine derivative (LQ8)

LQ4 and LQ8 showed better inhibition against kpUGM than against mtUGM. The basis of this selectivity is not clear. Unfortunately, inhibition of kpUGM by LQ4 and LQ8 were poor at lower concentrations. Nevertheless, the inhibition results of these two inhibitors suggest that it may be possible to design inhibitors specific to one particular UGM homologue.

5.6.3. Amino-sugar

The amino-sugar, SMC343 showed better inhibition against kpUGM, but poor inhibition at lower concentrations. The lack of groups that could interact at the diphosphate binding region and the uridine binding region may account for the poor inhibition at lower concentrations. Sugar-mimetics like SMC343 may occupy the sugar binding region where the majority of the interactions are through water molecules and not through direct interaction with the amino acid residues. The potency of these inhibitors could be optimized by incorporating groups that can interact with the residues at the diphosphate and uridine binding regions.

5.6.4. Uridine analog

ASK5 is a bis-uridine analog, which could potentially occupy the uridine-binding region of the active site. This compound showed poor inhibition against kpUGM and moderate inhibition against mtUGM. The presence of two uridine units in this compound may have some steric clashes at the sugar-binding region that could affect its inhibitory activity. **5.6.5. Bridged-sugars and sugar phosphates**

The bridged-sugar analogs (ASK1 and ASK2) and sugar phosphate (ASK4) showed poor inhibition against kpUGM and mtUGM. The poor inhibitory activity of

these compounds could be due to the lack of structural features that can bind at the diphosphate and uridine binding region. The majority of the interactions for the natural substrate (UDP-Gal*p*) are mediated by the active site residues of the uridine and diphosphate binding regions. The poor inhibition of bridged-sugars (ASK1 and ASK2) and the sugar phosphate (ASK4) are consistent with the previous inhibition studies of various sugar analogs (Richards & Lowary, 2009).

5.6.6. Substituted sulfonamides

Among the substituted sulfonamide derivatives, RIR286 and RIR299 showed the best inhibition against kpUGM and mtUGM. The presence of additional hydrogen bond donor/acceptor groups and the hydrophobic groups in RIR286 and RIR299 may contribute to their improved inhibition compared to RIR163. The positive charge on the amino group of RIR163 may have electrostatic repulsion with the active site arginines and this could result in poor binding and inhibition. RIR299 showed similar inhibition against kpUGM and mtUGM, but RIR286 showed better inhibition against kpUGM. This suggests that substitution of different hydrophobic groups in this class of inhibitors can alter their specificity.

5.7 Inhibitory activity towards drUGM WT

In order to use the drUGM:substrate complex structure as a platform for inhibitor design, these inhibitors were evaluated for the activity towards drUGM. Inhibition studies were done in a manner similar to the studies against kpUGM and mtUGM. Based on the overall structural similarity and conserved active site residues between drUGM and other bacterial UGMs (Table 4-1), we anticipated that the inhibition profile towards drUGM would be similar to kpUGM and mtUGM. But, most of the inhibitors showed poor

inhibition (at 1 mM) towards drUGM (Table 5-6). In order to understand this behavior, I did a careful analysis of the active site interactions in $drUGM_{ox}$ and identified an active site asparagine that may play a role in inhibitor binding.

Table 5-6 Inhibition of against drUGM WT and drUGM N372D.

Compound	% inhibition
	drUGM WT
LQ1	15 (± 5)*
LQ6	20 (± 2)
LQ4	44 (± 3)
LQ8	25 (± 5)
LQ10	20 (± 5)
SMC343	27 (± 2)
RIR163	37 (± 4)
RIR299	38 (± 2)
RIR286	42 (± 2)
ASK1	48 (± 3)
ASK2	31 (± 3)
ASK4	30 (± 3)
ASK5	58 (± 5)

* The errors represent the standard deviation based on duplicate measurements

5.8 Active site comparison and design of drUGM N372D

The active site comparison between drUGM, kpUGM and mtUGM revealed a minor difference with respect to one residue (Table 4-1). A conserved aspartic acid residue (Asp368 in mtUGM and Asp351 in kpUGM) is replaced by an asparagine residue (Asn372) in drUGM. Analysis of the drUGM:substrate complex structure revealed that

the Asn372 is involved in hydrogen bond interactions with the sugar moiety of the substrate. I proposed that this residue might be involved in inhibitor recognition. To test this hypothesis, I mutated the active site Asn372 to aspartic acid (N372D). drUGM N372D mutant was prepared by PCR-based site-directed mutagenesis (See section 3.0 and 3.1). Sequencing results confirmed the presence of desired mutation (N372D).

5.9 Kinetic characterization of drUGM N372D mutant

In order to determine the kinetic behavior of drUGM N372D towards the substrate, kinetic studies on drUGM N372D were carried out and compared the kinetic parameters with drUGM WT. The saturation curve and the kinetic parameters are shown in Figure 5-9 and Table 5-7 respectively.

The k_{cat} of drUGM N372D is comparable to drUGM WT, however, K_m increased by two-fold. The k_{cat}/K_m of drUGM WT and N372D mutant are comparable and it can be

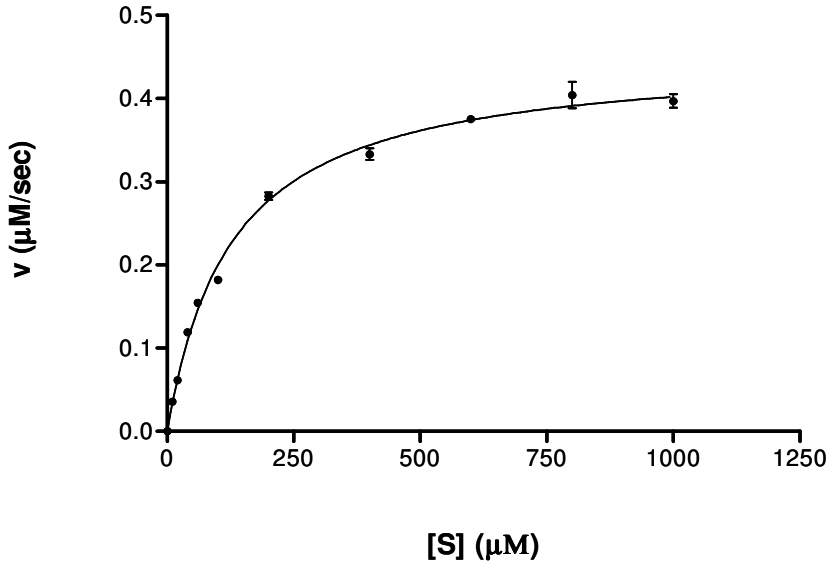


Figure 5-9 Saturatuion curve for drUGM N372D mutant.

concluded that this mutation does not affect the catalytic property of the enzyme towards the substrate.

Table 5-7 Kinetic parameters for drUGM N372D.

Enzyme	$k_{cat}(s^{-1})$	K_m (μM)	k_{cat}/K_m (μM^{-1} s^{-1})
drUGM N372D	45 ± 0.7	126	0.4
drUGM WT	66 ± 2	55 ± 7	1.2

5.10 Structure of drUGM N372D:UDP-Gal*p* complex

Our next goal was to determine if the drUGM N372D mutant binds the substrate (UDP-Gal*p*) similar to drUGM WT. drUGM N372D mutant was cocrystallized with UDP-Gal*p* using the same conditions that gave drUGM:UDP-Gal*p* complex crystals. The drUGM N372D:UDP-Gal*p* complex crystallized in the $P2_12_12_1$ space group and the crystals diffracted to 2.5 Å resolution. The space group and unit cell dimensions of drUGM N372D:UDP-Gal*p* crystals were similar to $drUGM_{ox}$ crystals. The data collection and refinement statistics for the drUGM N372D:UDP-Gal*p* complex are summarized in Table 5-8. The drUGM N372D:UDP-Gal*p* complex was solved using $drUGM_{ox}$ (without ligands) as the starting model for refinement. The simulated annealing difference map revealed electron density for FAD and the nucleotide portion (UDP) of the substrate. FAD and UDP were modeled into the density and further rounds of refinement were carried out (Figure 5-10). The density for sugar moiety was weak in all the chains and this could be due to the poor occupancy of the substrate. This is consistent with the $drUGM_{ox}$ structure, where the density for sugar moiety was poor when compared to $drUGM_{red}$ (See section 4.2) The NCS average map over 10 active sites showed better density for the sugar portion of the substrate in the drUGM N372D:UDP-Gal*p* complex.

The overall structure of drUGM N372D:UDP-Gal*p* is similar to $drUGM_{ox}$ (rmsd of 0.6 Å for all equivalent C_α atoms) including the two mobile loops (Figure 5-9). Most importantly, the binding mode and active site interactions of UDP-Gal*p* in drUGM N372D were similar to drUGM WT. The structure of the drUGM N372D:UDP-Gal*p* complex suggests that this mutation did not perturb the substrate binding mode and its interactions at the active site.

Table 5-8 Data collection statistics and refinement.

	drUGM N372D:UDP-Gal*p*
Data collection	
Space group	$P2_12_12_1$
Cell dimensions	
a, *b*, *c* (Å)	133.0, 176.4, 220.7 90.00, 90.00, 90.00
Resolution (Å)	2.50 (2.59-2.50)
R_{sym}	18.9 (61.6)
$I / \sigma I$	4.7 (1.5)
Completeness (%)	100 (99.8)
Redundancy	7.3
Refinement	
Resolution (Å)	40.0-2.5
No. reflections	179222
R_{work} / R_{free}	0.22/0.27
No. atoms	
Protein	30134
Ligand (FAD, UDP and UDP-Gal*p*)	857
Water (molecules)	401
B-factors	
Protein	53.1
Ligand	59.5 (UDP-Gal*p*) 61.8 (UDP) 54.7 (FAD)
Water	41.5
r.m.s. deviations	
Bond lengths (Å)	0.008
Bond angles (°)	1.2

*values in the parenthesis are for the highest resolution shell

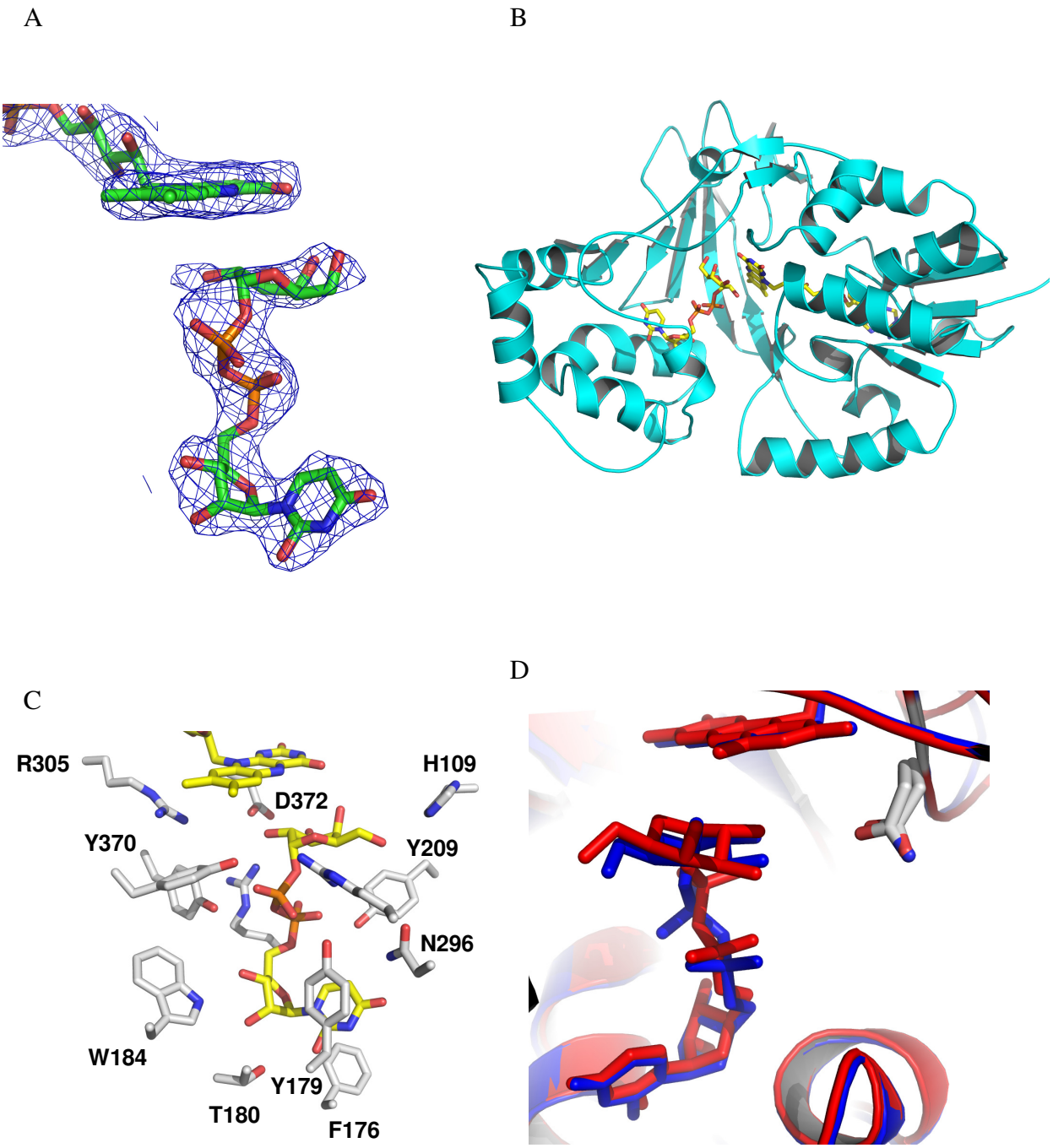


Figure 5-10 Structure of drUGM N372D:UDP-Gal*p* complex. **A)** $2F_o$-F_c map of UDP-Gal*p* and FAD. The density of UDP-Gal*p* is weak and similar to that observed in the $drUGM_{ox}$ complex. **B)** The overall structure, mobile loop conformation and the binding mode of UDP-Gal*p* in drUGM N372D. **C)** Active site interactions of UDP-Gal*p* in drUGM N372D mutant. **D)** Overlay of $drUGM_{ox}$ (blue) and drUGM N372:UDP-Gal*p* (red) to highlight the binding mode similarity of in both the complexes.

5.11 Inhibition profile against drUGM N372D mutant

Kinetic and structural studies of drUGM N372D mutant confirmed that this mutant behaves in a similar way as drUGM WT towards the substrate. Inhibition studies were performed with the drUGM N372D to study its behavior towards the inhibitors (Table 5-9). It is evident from Table 5-9 that most of the inhibitors showed better inhibition towards the drUGM N372D as compared to drUGM WT. Some inhibitors showed better activity even at low concentrations. Comparison of the inhibition studies between drUGM WT and drUGM N372D mutant suggests that the active site aspartic acid plays a role in binding to these inhibitors. However, the structural basis for the observed differences in the inhibition profile towards drUGM N372D mutant is unclear at this moment.

5.12 Inhibition against kpUGM D351N mutant

The inhibition profile towards the drUGM N372D mutant revealed the role of an active site aspartic acid in inhibitor binding. It is possible that mutation of this active site aspartic acid in kpUGM or mtUGM to asparagine may affect the inhibitor binding and decrease the % inhibition. kpUGM D351N mutant was designed and prepared by PCR-based site-directed mutageneis method (See section 3.0 and 3.1). The inhibitory activity of one of the promising inhibitors (LQ10) was evaluated with kpUGM D351N. LQ10 showed only 8 % inhibition at 1.0 µM towards kpUGM D351N, while approximately 50 % inhibition was observed with kpUGM WT (Table 5-9). This suggests that the active aspartic acid residue has a definite role in determining inhibitor specificity.

5.13 Role of active site aspartic acid

The mutation of active site Asn to Asp (N372D in drUGM) and Asp to Asn

Table 5-9 Comparion of inhibitory profile between drUGM WT and drUGM N372D.

Compound	% inhibition	
	drUGM WT	drUGM N372D
LQ1	15 (± 5)*	56 (± 1) @10 μM
LQ6	20 (± 2)	60 (± 3) @ 1.0 μM
LQ4	44 (± 3)	68 (± 6)
LQ8	25 (± 7)	49 (± 2)
LQ10	20 (± 5)	76 (± 2) @ 1 μM
SMC343	27 (± 2)	36 (± 2)
RIR163	37 (± 4)	33 (± 5)
RIR299	38 (± 2)	77 (± 4)
RIR286	42 (± 2)	37 (± 4)
ASK1	48 (± 3)	52 (± 3)
ASK2	31 (± 3)	43 (± 4)
ASK4	30 (± 3)	38 (± 3)
ASK5	58 (± 5)	52 (± 8

-- not determined
* The errors represent the standard deviation based on duplicate measurements

Table 5-10 Inhibition of LQ10 towards kpUGM WT and D351N mutant.

	% inhibition	
	kpUGM WT	kpUGM D351N
LQ10	48 (± 1)	8 (± 2)

* The errors represent the standard deviation based on duplicate measurements

(D351N in kpUGM) showed altered specificity towards inhibitors. From a structural standpoint, Asn and Asp have similar geometry and can accept/donate hydrogen bonds,

but Asp being a carboxylic acid can exist in ionized or unionized states in the active site. The ionized state of Asp may participate in strong ionic interactions with the inhibitor and improve its binding (as observed with drUGM N372D mutant). The ionized form of Asp may be stabilized by an arginine or lysine (located nearby) and lock the conformation of Asp. To understand the behavior of the Asn to Asp mutation, I looked into the literature for similar mutations and their behavior towards the substrate and inhibitor. Indeed, Asn→Asp and Asp→Asn mutants showed significant differences in activity and inhibition. Asp→Asn mutation (D27N) in dihydrofolate reductase (DHFR) was carried out to probe the binding of folate analogs, such as methotrexate (MTX) (Villafranca et al, 1986). The K_d for MTX towards the D27N increased to 1.9 nM from 0.07 nM. Structures of WT DHFR:MTX complex and DHFR D27N:MTX complexes were determined and the inhibitor binds in an identical orientation in both WT and the D27N mutant, although the positions of two active site water molecules showed some differences.

Mutation of an active site Asp→Asn in phosphodiesterase 4D3 (PDE4D3) altered the substrate specificity and inhibition profile (Herman et al, 2000). PDE4D3 catalyzes the hydrolysis of cyclic AMP (cAMP), a second messenger involved in the cell signaling pathways. Asp413 was mutated to Asn (D413N) to probe its role in catalysis and inhibitor binding. The K_m for D413N mutant was found to be 16 μM (5.8 μM for WT). It is reasonable to assume that this mutation did not affect the substrate binding significantly. Most importantly, the D413N mutant was shown to hydrolyze cGMP. A direct comparison of kinetic results for the hydrolysis of cGMP by WT and the D413N mutant was hampered due to the inability to measure hydrolysis of cGMP by the WT

(levels of hydrolysis product were low for WT). The D413N mutant showed resistance or a low level of inhibition towards rolipram, a known inhibitor of PDE4. The IC_{50} value of rolipram is 1.0 μM against WT and for the D413N mutant it was found to be 34 μM. The structural basis for the altered substrate specificity and inhibitor resistance by the D413N mutant is lacking.

In another study, mutation of an active site asparagine to aspartic acid altered the substrate specificity significantly (Liu & Santi, 1992). The enzyme thymidylate synthase (TS) catalyzes the reductive methylation of dUMP by tetrahydrofolate (THF) to form dTMP and dihydrofolate (DHF). To probe the role of active site asparagine (N229), it was mutated to aspartic acid (N229D). The K_m of dUMP and THF for N229D mutant increased by 3.5 and 10-fold respectively, while k_{cat} decreased by 1000-fold. Nonetheless, this mutation altered the substrate specificity from dUMP to dCMP. The N229D mutant showed 40-fold higher specificity for dCMP than for dUMP. dCMP is not a substrate for TS (WT) and binds very poorly.

The behavior of the drUGM N372D and kpGM D351N mutants towards the inhibitor is not surprising. Precedence for such effects observed in other proteins clearly assigns a specific role for the active site aspartic acid in UGM towards inhibitor binding. Further structural studies are needed to understand the role of this active site aspartic acid.

CHAPTER 6

Conclusions

6.0 Conclusions and Future Directions

Structural and inhibition studies on UGM provided valuable information regarding the catalytic mechanism and inhibition on this unique enzyme. Our objective of using alternate homolog to solve UGM:substrate complex structure was achieved successfully through drUGM. Also, I was able to use drUGM for solving other ligand complexes. Structures of drUGM in complex with the substrate (UDP-Gal*p*) gave insights into the substrate binding modes and conformational changes associated with the substrate binding (Chapter 4). The binding mode of UDP-Gal*p* to UGM is unique and not observed in other UDP-Gal*p* binding proteins. The structure of $drUGM_{red}$ provided a structural basis for the role of FAD in the reaction. The role of mobile loops in stabilizing the bound substrate is evident from the drUGM:substrate complex structures. The role of two conserved arginines (Arg198 and Arg305) and the tyrosines are evident from the drUGM:substrate complex structures and is consistent with the mutagenesis studies on kpUGM and ecUGM. In Chapter 4, I also discuss the comparison of the substrate complex structures from drUGM with kpUGM. Significant differences in the binding mode and mobile loop conformations were observed between the oxidized substrate complexes from kpUGM and drUGM. Base on these comparisons, it appears that the drUGM:substrate complex structures reported here are a closer approximation of productive substrate binding when compared to kpUGM:substrate complex structure.

In Chapter 4, I discuss how UDP, UMP and UDP-Glc bind to drUGM. The lack of the sugar moiety in UDP does not seem to affect the mobile loop movement and exists

in a closed conformation similar to drUGM:substrate complex structures. The drUGM:UMP structures showed that the absence of β-phosphate group in UDP does not affect the mobile loop closure. The binding of UDP-Glc (discussed in Chapter 4) underlines the role and orientation of the C4 hydroxyl group of the sugar moiety and its effect on the binding mode and active site interactions. A change from C4 hydoxyl axial (UDP-Gal*p*) to C4 hydoxyl equatorial (UDP-Glc) dramatically altered the binding orientation of the sugar moiety (UDP-Glc) and the active site interactions. The binding mode of UDP-Glc in kpUGM and drUGM is similar, but differences in the mobile loop conformation and stabilization by active site arginines were observed. In general, the mobile loops in drUGM:ligand complexes adopt a fully closed conformation across all the chains, however the mobile loops adopt different conformational states in the kpUGM:ligand complexes. Differences in the energetics of loop closure between drUGM and kpUGM may account for different conformational states of the mobile loop, however the basis for this difference is unclear.

In Chapter 5, results from the design and screening of inhibitors from two different approaches, ligand-based and structure-based methods were discussed. The activity of GCP, the phosphonate analog of UDP-Gal*p* (ligand-based design) was screened against various bacterial UGMs and showed only weak inhibition. The unique binding mode and loss of stabilization by the active site arginines (Arg198 and Arg305) account for the weak activity of GCP. The drUGM:GCP complex structure underlines the consequence of substrate guided design (phosphonate analogs) of inhibitors for UGM. Although phosphonate analogs of natural substrates have proven to be successful for other enzymes, UGM seems to be an exception. In Chapter 5, I also discuss the

identification of some novel inhibitors of UGM through SBVS. The promising inhibitors showed activity close to 1.0 μM and share a common indole scaffold. Docking studies allowed us to predict the binding mode of these indole-based inhibitors (LQ1, LQ6 and LQ10). The indole nucleus showed stacking interactions with the Tyr179 and Phe176, an interaction observed with the uracil ring of the substrate. The hydrazine moiety and the tail portion of the inhibitor interact with other active site residues. The indole-based inhibitors are promising candidates or lead molecules for further optimization of potency.

The closed conformation of the mobile loop is considered as a productive conformation for UGM and hence docking of inhibitors were done using the closed conformation of the mobile loop. However, one cannot rule out the possibility of inhibitors binding to UGM with open conformation of the mobile loop. Precedence for such inhibitors for UGM is currently not known. Crystal structures UGM in complex with the indole-based inhibitors (LQ6 or LQ10) may provide information on the conformational state of the mobile loop (open *vs* closed) and can further assist in the inhibitor design.

Inhibition studies with drUGM WT and drUGM N372D mutant revealed the role of an active site aspartic acid in inhibitor binding. Muation of N372 → Asp (drUGM N372D) did not affect substrate binding and catalytic activity significantly. Most of the inhibitors showed better activity towards drUGM N372D mutant as compared to drUGM WT suggestive of the importance of this residue. To probe the role of this aspartic acid in kpUGM, I mutated it to asparagine (kpUGM D351N) and screened against LQ10. The inhibitory activity of LQ10 towards kpUGM D351N mutant dropped significantly as compared to the kpUGM WT. The inhibition studies with drUGM N372D and kpUGM

D351N assign a role for active site aspartic acid, although the structural basis of its role in inhibitor binding is unclear.

The structural and enzyme inhibition studies on UGM reported here have opened new questions or directions for further investigations.

6.1 Kinetic characterization of drUGM active site mutants

Based on the drUGM:UDP-Gal*p* complex structures, other active site residues were identified that are involved in the substrate binding. The role of these active site residues have not been studied before in any UGM. Notably, Arg364 located in the FAD binding domain has moved towards the substrate binding pocket, whereas, in unliganded structures (kpUGM, ecUGM and mtUGM) the corresponding arginine is interacting with the phosphate group of the FAD. Towards this goal (to understand the role of other active site amino acids in substrate binding and catalysis), I have prepared these mutants by PCR-based site-directed mutagenesis and sequencing studies confirmed the desired mutation. The clones were transformed into *E. coli* expression strain and are ready for expression, purification and kinetic characterization. These active site-mutants would provide insights onto the role of these active site residues in substrate binding and catalysis.

6.2 Crystallization of unliganded drUGM

Until now, I am unable to determine the structure of unliganded drUGM. This structure would provide valuable information on the conformational changes in comparison to the drUGM:substrate complex. For example the mobile loop conformation and the side chain orientations of some active site residues (Arg364 and Tyr370). Crystals of unliganded drUGM diffracted to only 5.0 Å and hence further optimization to

improve the diffraction is required. Alternatively, the active site mutants of drUGM can be used to crystallize unliganded drUGM.

6.3 Crystallization of UGM-inhibitor complexes

Inhibition studies on various bacterial UGMs led to the identification of some novel indole-based inhibitors. To understand the binding mode and active site interactions of these inhibitors, crystal structures of UGM-inhibitor complexes needs to determined. The choice of UGM for structural investigations could be the drUGM N372D mutant as it showed activity comparable to the kpUGM and mtUGM. Moreover, I have shown in Chapter 4 that drUGM can be used as a model for understanding bacterial UGM. In addition, drUGM:ligand complexes crystallized easily relative to other bacterial UGMs; however, one could also attempt to crystallize these inhibitors with kpUGM. Sufficient quantity of LQ6 (indole-based inhibitor) is available for cocrystallization or soaking. Initial attempts to crystallize the drUGM N372D:LQ6 complex gave crystals (using conditions similar to drUGM:substrate) that diffracted to only 4.5 Å. The structure of the UGM:LQ6 complex would provide valuable information on inhibitior binding, notably the role of aspartic acid and should assist in the structure-guided optimization of these inhibitors.

REFERENCES

Adams P, Afonine P, Bunkóczi G, Chen V, Davis I, Echols N, Headd J, Hung L, Kapral G, Grosse-Kunstleve R, McCoy A, Moriarty N, Oeffner R, Read R, Richardson D, Richardson J, Terwilliger T, Zwart P (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66:** 213-221

Adams P, Grosse-Kunstleve R, Hung L, Ioerger T, McCoy A, Moriarty N, Read R, Sacchettini J, Sauter N, Terwilliger T (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* **58:** 1948-1954

Aronov A, Munagala N, Ortiz De Montellano P, Kuntz I, Wang C (2000) Rational design of selective submicromolar inhibitors of Tritrichomonas foetus hypoxanthine-guanine-xanthine phosphoribosyltransferase. *Biochemistry* **39:** 4684-4691

Barlow J, Blanchard J (2000) Enzymatic synthesis of UDP-(3-deoxy-3-fluoro)-D-galactose and UDP-(2-deoxy-2-fluoro)-D-galactose and substrate activity with UDP-galactopyranose mutase. *Carbohydr Res* **328:** 473-480

Barlow J, Girvin M, Blanchard J (1999) Positional isotope exchange catalyzed by UDP-galactopyranose mutase. *J Am Chem Soc* **121:** 6968-6969

Barton G (2008) Sequence alignment for molecular replacement. *Acta Crystallogr D Biol Crystallogr* **64:** 25-32

Baxter C, Murray C, Waszkowycz B, Li J, Sykes R, Bone R, Perkins T, Wylie W (2000) New approach to molecular docking and its application to virtual screening of chemical databases. *J Chem Inf Comput Sci* **40:** 254-262

Beis K, Srikannathasan V, Liu H, Fullerton S, Bamford V, Sanders D, Whitfield C, McNeil M, Naismith J (2005) Crystal structures of Mycobacteria tuberculosis and Klebsiella pneumoniae UDP-galactopyranose mutase in the oxidised state and Klebsiella pneumoniae UDP-galactopyranose mutase in the (active) reduced state. *J Mol Biol* **348:** 971-982

Berman H, Westbrook J, Feng Z, Gilliland G, Bhat T, Weissig H, Shindyalov I, Bourne P (2000) The Protein Data Bank *Nucleic Acids Research* 28: 235-242

Binda C, Coda A, Angelini R, Federico R, Ascenzi P, Mattevi A (1999) A 30-angstrom-long U-shaped catalytic tunnel in the crystal structure of polyamine oxidase. *Structure* **7:** 265-276

Bissantz C, Folkers G, Rognan D (2000) Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J Med Chem* **43:** 4759-4767

Bleile, D.W. (2008). Structural Models of UDP-galactopyranose Mutase-Ligand Binding by Interpretation of Saturation Transfer Difference NMR Spectra through Computational Modeling. MSc Thesis, Simon Fraser University

Boehm H, Boehringer M, Bur D, Gmuender H, Huber W, Klaus W, Kostrewa D, Kuehne H, Luebbers T, Meunier-Keller N, Mueller F (2000) Novel inhibitors of DNA gyrase: 3D structure based biased needle screening, hit validation by biophysical methods, and 3D guided optimization. A promising alternative to random screening. *J Med Chem* **43:** 2664-2674

Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**:248-254

Burton A, Wyatt P, Boons G (1997) Preparation of fluorinated galactosyl nucleoside diphosphates to study the mechanism of the enzyme galactopyranose mutase. *J Chem Soc Perkin Trans* **1:** 2375-2382

Campbell J, Duée E, Hodgson G, Mercer W, Stammers D, Wendell P, Muirhead H, Watson H (1972) X-ray diffraction studies on enzymes in the glycolytic pathway. *Cold Spring Harb Symp Quant Biol* **36:** 165-170

Caravano A, Mengin-Lecreulx D, Brondello J, Vincent S, Sinaÿ P (2003) Synthesis and inhibition properties of conformational probes for the mutase-catalyzed UDP-galactopyranose/furanose interconversion. *Chemistry* **9:** 5888-5898

Caravano A, Vincent S (2009) Synthesis of Three C-Glycoside Analogues of UDP-Galactopyranose as Conformational Probes for the Mutase-Catalyzed Furanose/Pyranose Interconversion. *Eur J Org Chem* 1771-1780

Carlson E, May J, Kiessling L (2006) Chemical probes of UDP-galactopyranose mutase. *Chem Biol* **13:** 825-837

Chad J, Sarathy K, Gruber T, Addala E, Kiessling L, Sanders D (2007) Site-directed mutagenesis of UDP-galactopyranose mutase reveals a critical role for the active-site, conserved arginine residues. *Biochemistry* **46:** 6723-6732

Chayen N (1997) The role of oil in macromolecular crystallization. *Structure* **5:** 1269-1274

Chayen N (1998) Comparative studies of protein crystallization by vapour-diffusion and microbatch techniques. *Acta Crystallogr D Biol Crystallogr* **54:** 8-15

Cho Y, Ioerger T, Sacchettini J (2008) Discovery of novel nitrobenzothiazole inhibitors for Mycobacterium tuberculosis ATP phosphoribosyl transferase (HisG) through virtual screening. *J Med Chem* **51:** 5984-5992

Clarke A, Hurtado-Guerrero R, Pathak S, Schuttelkopf A, Borodkin V, Shepherd S, Ibrahim A, van Aalten D (2008) Structural insights into mechanism and specificity of O-GlcNAc transferase. *EMBO J* **27:** 2780-2788

Corradi H, Corrigall A, Boix E, Mohan C, Sturrock E, Meissner P, Acharya K (2006) Crystal structure of protoporphyrinogen oxidase from Myxococcus xanthus and its complex with the inhibitor acifluorfen. *J Biol Chem* **281:** 38625-38633

Cummings M, DesJarlais R, Gibbs A, Mohan V, Jaeger E (2005) Comparison of automated docking programs as virtual screening tools. *J Med Chem* **48:** 962-976

Dauter Z (1999) Data-collection strategies. *Acta Crystallogr D Biol Crystallogr* **55:** 1703-1717

Davis I, Murray L, Richardson J, Richardson D (2004) MolProbity: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Research* **32**: W615-W619

De Azevedo WJ (2010) Structure-based virtual screening. *Curr Drug Targets* **11:** 261-263

de Lederkremer R, Casal O, Alves M, Colli W (1980) Evidence for the presence of D-galactofuranose in the lipopeptidophosphoglycan from Trypanosome cruzi. Modification and tritium labeling. *FEBS Lett* **116:** 25-29

Derewenda Z (2004) The use of recombinant methods and molecular engineering in protein crystallization. *Methods* **34:** 354-363

Desvergnes S, Desvergnes V, Martin O, Itoh K, Liu H, Py S (2007) Stereoselective synthesis of beta-1-C-substituted 1,4-dideoxy-1,4-imino-D-galactitols and evaluation as UDP-galactopyranose mutase inhibitors. *Bioorg Med Chem* **15:** 6443-6449

Doman T, McGovern S, Witherbee B, Kasten T, Kurumbail R, Stallings W, Connolly D, Shoichet B (2002) Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *J Med Chem* **45:** 2213-2221

Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60:** 2126-2132

Engle R (1977) Phosphonates as analogs of natural phosphates. *Chem Rev* **77:** 349-367

Errey J, Mann M, Fairhurst S, Hill L, McNeil M, Naismith J, Percy J, Whitfield C, Field R (2009) Sugar nucleotide recognition by Klebsiella pneumoniae UDP-D-galactopyranose mutase: fluorinated substrates, kinetics and equilibria. *Org Biomol Chem* **7:** 1009-1016

Evans P, McCoy A (2008) An introduction to molecular replacement. *Acta Crystallogr D Biol Crystallogr* **64:** 1-10

Filikov A, Mohan V, Vickers T, Griffey R, Cook P, Abagyan R, James T (2000) Identification of ligands for RNA targets via structure-based virtual screening: HIV-1 TAR. *J Comput Aided Mol Des* **14:** 593-610

Fullerton S, Daff S, Sanders D, Ingledew W, Whitfield C, Chapman S, Naismith J (2003) Potentiometric analysis of UDP-galactopyranose mutase: stabilization of the flavosemiquinone by substrate. *Biochemistry* **42:** 2104-2109

Ghavami A, Chen J, Mario Pinto B (2004) Synthesis of a novel class of sulfonium ions as potential inhibitors of UDP-galactopyranose mutase. *Carbohydr Res* **339:** 401-407

Ghosh S, Nie A, An J, Huang Z (2006) Structure-based virtual screening of chemical libraries for drug discovery. *Curr Opin Chem Biol* **10:** 194-202

Goodsell D, Morris G, Olson A (1996) Automated docking of flexible ligands: applications of AutoDock. *J Mol Recognit* **9:** 1-5

Gordon R, Sivarajah P, Satkunarajah M, Ma D, Tarling C, Vizitiu D, Withers S, Rini J (2006) X-ray crystal structures of rabbit N-acetylglucosaminyltransferase I (GnT I) in complex with donor substrate analogues. *J Mol Biol* **360:** 67-79

Gouet P, Courcelle E, Stuart D, Métoz F (1999) ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15:** 305-308

Gouet P, Robert X, Courcelle E (2003) ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res* **31:** 3320-3323

Green D (2002) The bacterial cell wall as a source of antibacterial targets. *Expert Opin Ther Targets* **6:** 1-19

Gruber T, Borrok M, Westler W, Forest K, Kiessling L (2009a) Ligand binding and substrate discrimination by UDP-galactopyranose mutase. *J Mol Biol* **391:** 327-340

Gruber T, Westler W, Kiessling L, Forest K (2009b) X-ray crystallography reveals a reduced substrate complex of UDP-galactopyranose mutase poised for covalent catalysis by flavin. *Biochemistry* **48:** 9171-9173

Grüneberg S, Stubbs M, Klebe G (2002) Successful virtual screening for novel inhibitors of human carbonic anhydrase: strategy and experimental confirmation. *J Med Chem* **45:** 3588-3602

Hajduch J, Nam G, Kim E, Frohlich R, Hanover J, Kirk K (2008) A convenient synthesis of the C-1-phosphonate analogue of UDP-GlcNAc and its evaluation as an inhibitor of O-linked GlcNAc transferase (OGT). *Carbohydr Res* **343:**189-195

Hassell A, An G, Bledsoe R, Bynum J, Carter H, Deng S, Gampe R, Grisard T, Madauss K, Nolte R, Rocque W, Wang L, Weaver K, Williams S, Wisely G, Xu R, Shewchuk L (2007) Crystallization of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr* **63:** 72-79

Haynes C, Koder R, Miller A, Rodgers D (2002) Structures of nitroreductase in three states: effects of inhibitor binding and reduction. *J Biol Chem* **277:** 11513-11520

Herman S, Juilfs D, Fauman E, Juneau P, Menetski J (2000) Analysis of a mutation in phosphodiesterase type 4 that alters both inhibitor activity and nucleotide selectivity. *Mol Pharmacol* **57:** 991-999

Hopkins S, Vale R, Kuntz I (2000) Inhibitors of kinesin activity from structure-based computer screening. *Biochemistry* **39:** 2805-2814

Huang Z, Zhang Q, Liu H (2003) Reconstitution of UDP-galactopyranose mutase with 1-deaza-FAD and 5-deaza-FAD: analysis and mechanistic implications. *Bioorg Chem* **31:** 494-502

Itoh K, Huang Z, Liu H (2007) Synthesis and analysis of substrate analogues for UDP-galactopyranose mutase: implication for an oxocarbenium ion intermediate in the catalytic mechanism. *Org Lett* **9:** 879-882

Jain A (2003) Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *J Med Chem* **46:** 499-511

Joiner K (1988) Complement evasion by bacteria and parasites. *Annu Rev Microbiol* **42:** 201-230

John M (2007) Enzyme activity: reversible inhibition. Handbook of Proteins, **1:** 482-490

Kabsch W (2010a) Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr* **66:** 133-144

Kabsch W (2010b) XDS. *Acta Crystallogr D Biol Crystallogr* **66:** 125-132

Karunan Partha S, Bonderoff S, van Straaten K, Sanders D (2009) Expression, purification and preliminary X-ray crystallographic analysis of UDP-galactopyranose

mutase from Deinococcus radiodurans. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **65:** 843-845

Keegan R, Winn M (2008) MrBUMP: an automated pipeline for molecular replacement. *Acta Crystallogr D Biol Crystallogr* **64:** 119-124

Kellenberger E, Rodrigo J, Muller P, Rognan D (2004) Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins* **57:** 225-242

Kiss R, Kiss B, Könczöl A, Szalai F, Jelinek I, László V, Noszál B, Falus A, Keseru G (2008) Discovery of novel human histamine H4 receptor ligands by large-scale structure-based virtual screening. *J Med Chem* **51:** 3145-3153

Knirel Y, Kochetkov N (1994) The structure of lipopolysaccharides of Gram-negative bacteria 3. The structure of O-antigens - A Review. *Biochemistry (Moscow)* **59:** 1325-1383

Kramer B, Rarey M, Lengauer T (1999) Evaluation of the FlexX incremental construction algorithm for protein-ligand docking. *Proteins* **37:** 228-241

Kraut J (1988) How do enzymes work? *Science* **242:** 533-540

Kremer L, Besra G (2002) Current status and future development of antitubercular chemotherapy. *Expert Opin Investig Drugs* **11:** 1033-1049

Köplin R, Brisson J, Whitfield C (1997) UDP-galactofuranose precursor required for formation of the lipopolysaccharide O antigen of Klebsiella pneumoniae serotype O1 is synthesized by the product of the rfbDKPO1 gene. *J Biol Chem* **272:** 4121-4128

Lee R, Smith M, Nash R, Griffiths R, McNeil M, Grewal R, Yan W, Besra G, Brennan P, Fleet G (1997) Inhibition of UDP-Gal mutase and mycobacterial galactan biosynthesis by pyrrolidine analogues of galactofuranose. *Tetrahed. Lett* **38:** 6733-6736

Lee R, Smith M, Pickering L, Fleet G (1999) An approach to combinatorial library generation of galactofuranose mimics as potential inhibitors of mycobacterial cell wall biosynthesis: Synthesis of a peptidomimetic of uridine 5 '-diphosphogalactofuranose (UDP-Galf). *Tetrahed. Lett* **40:** 8689-8692

Lennon B, Williams CJ, Ludwig M (1999) Crystal structure of reduced thioredoxin reductase from Escherichia coli: structural flexibility in the isoalloxazine ring of the flavin adenine dinucleotide cofactor. *Protein Sci* **8:** 2366-2379

Leslie A (2006) The integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr* **62:** 48-57

Liautard V, Christina A, Desvergnes V, Martin O (2006) Diastereoselective synthesis of novel iminosugar-containing UDP-Galf mimics: potential inhibitors of UDP-Gal mutase and UDP-Galf transferases. *J Org Chem* **71:** 7337-7345

Liu L, Santi D (1992) Mutation of asparahine-229 to aspartate in thymidylate synthase converts the enzyme to a deoxycytidylate methylase. *Biochemistry* **31:** 5100-5104

Lyne P (2002) Structure-based virtual screening: an overview. *Drug Discov Today* **7:** 1047-1055

Mansoorabadi S, Thibodeaux C, Liu H (2007) The diverse roles of flavin coenzymes--nature's most versatile thespians. *J Org Chem* **72:** 6329-6342

McPherson A (2004) Introduction to protein crystallization. *Methods* **34:** 254-265

McPherson A (2009) Introduction to the Crystallization of Biological Macromolecules. *Curr. Top. Memb* **63:** 5-23

Mestres J (2002) Virtual screening: a real screening complement to high-throughput screening. *Biochem Soc Trans* **30:** 797-799

Miguet L, Zervosen A, Gerards T, Pasha F, Luxen A, Distèche-Nguyen M, Thomas A (2009) Discovery of new inhibitors of resistant Streptococcus pneumoniae penicillin binding protein (PBP) 2x by structure-based virtual screening. *J Med Chem* **52:** 5926-5936

Moonen C, Vandenberg W, Boerjan M, Muller F (1984) C-13 AND N-15 nuclear magnetic-resonance study on the interaction between riboflavin and riboflavin-binding apoprotein *Biochemistry* **23:** 4873-4878

Murshudov G, Vagin A, Lebedev A, Wilson K, Dodson E (1999) Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr D Biol Crystallogr* **55:** 247-255

Nassau P, Martin S, Brown R, Weston A, Monsey D, McNeil M, Duncan K (1996) Galactofuranose biosynthesis in Escherichia coli K-12: identification and cloning of UDP-galactopyranose mutase. *J Bacteriol* **178:** 1047-1052

Pan F, Jackson M, Ma Y, McNeil M (2001) Cell wall core galactofuran synthesis is essential for growth of mycobacteria. *J Bacteriol* **183:** 3991-3998

Pan W, Ansiaux C, Vincent S (2007) Synthesis of acyclic galactitol- and lyxitol-aminophosphonates as inhibitors of UDP-galactopyranose mutase. *Tetrahed. Lett* **48:** 4353-4356

Partha S, van Straaten K, Sanders D (2009) Structural basis of substrate binding to UDP-galactopyranose mutase: crystal structures in the reduced and oxidized state complexed with UDP-galactopyranose and UDP. *J Mol Biol* **394:** 864-877

Pedersen L, Turco S (2003) Galactofuranose metabolism: a potential target for antimicrobial chemotherapy. *Cell Mol Life Sci* **60:** 259-266

Peltier P, Euzen R, Daniellou R, Nugier-Chauvin C, Ferrières V (2008) Recent knowledge and innovations related to hexofuranosides: structure, synthesis and applications. *Carbohydr Res* **343:** 1897-1923

Perola E, Xu K, Kollmeyer T, Kaufmann S, Prendergast F, Pang Y (2000) Successful virtual screening of a chemical database for farnesyltransferase inhibitor leads. *J Med Chem* **43:** 401-408

Persson K, Ly H, Dieckelmann M, Wakarchuk W, Withers S, Strynadka N (2001) Crystal structure of the retaining galactosyltransferase LgtC from Neisseria meningitidis in complex with donor and acceptor sugar analogs. *Nat Struct Biol* **8:** 166-175

Pflugrath J (1999) The finer things in X-ray diffraction data collection. *Acta Crystallogr D Biol Crystallogr* **55:** 1718-1725

Pflugrath J (2004) Macromolecular cryocrystallography--methods for cooling and mounting protein crystals at cryogenic temperatures. *Methods* **34:** 415-423

Ramsden N, Buetow L, Dawson A, Kemp L, Ulaganathan V, Brenk R, Klebe G, Hunter W (2009) A structure-based approach to ligand discovery for 2C-methyl-D-erythritol-2,4-cyclodiphosphate synthase: a target for antimicrobial therapy. *J Med Chem* **52:** 2531-2542

Rarey M, Kramer B, Lengauer T, Klebe G (1996) A fast flexible docking method using an incremental construction algorithm. *J Mol Biol* **261:** 470-489

Rhodes G (2006) Crystallography Made Crystal Clear. 3rd Ed, London, UK: Academic Press

Richards M, Lowary T (2009) Chemistry and biology of galactofuranose-containing polysaccharides. *Chembiochem* **10:** 1920-1938

Romanenko V, Kukhar V (2006) Fluorinated phosphonates: Synthesis and biomedical application. *Chem Rev* **106:** 3868-3935

Rose N, Completo G, Lin S, McNeil M, Palcic M, Lowary T (2006) Expression, purification, and characterization of a galactofuranosyltransferase involved in Mycobacterium tuberculosis arabinogalactan biosynthesis. *J Am Chem Soc* **128:** 6721-6729

Rose N, Zheng R, Pearcey J, Zhou R, Completo G, Lowary T (2008) Development of a coupled spectrophotometric assay for GlfT2, a bifunctional mycobacterial galactofuranosyltransferase. *Carbohydr Res* **343:** 2130-2139

Rossmann M (2001) Molecular replacement--historical background. *Acta Crystallogr D Biol Crystallogr* **57:** 1360-1366

Rossmann M, van Beek C (1999) Data processing. *Acta Crystallogr D Biol Crystallogr* **55:** 1631-1640

Rupp B (2010) Biomolecular Crystallography: Principles, Practice and Application to Structural Biology. 1st Ed, Garland Science, Taylor & Francis Group

Sanders D, Staines A, McMahon S, McNeil M, Whitfield C, Naismith J (2001) UDP-galactopyranose mutase has a novel structure and mechanism. *Nat Struct Biol* **8:** 858-863

Schapira M, Raaka B, Samuels H, Abagyan R (2000) Rational discovery of novel nuclear hormone receptor antagonists. *Proc Natl Acad Sci U S A* **97:** 1008-1013

Scherman M, Winans K, Stern R, Jones V, Bertozzi C, McNeil M (2003) Drug targeting Mycobacterium tuberculosis cell wall synthesis: development of a microtiter plate-based screen for UDP-galactopyranose mutase and identification of an inhibitor from a uridine-based library. *Antimicrob Agents Chemother* **47:** 378-382

Schneider G, Böhm H (2002) Virtual screening and fast automated docking methods. *Drug Discov Today* **7:** 64-70

Schulz-Gasch T, Stahl M (2003) Binding site characteristics in structure-based virtual screening: evaluation of current docking tools. *J Mol Model* **9:** 47-57

Seifert M (2009) Targeted scoring functions for virtual screening. *Drug Discov Today* **14:** 562-569

Smith R, Hubbard R, Gschwend D, Leach A, Good A (2003) Analysis and optimization of structure-based virtual screening protocols. (3). New methods and old problems in scoring function design. *J Mol Graph Model* **22:** 41-53

Soltero-Higgin M, Carlson E, Gruber T, Kiessling L (2004) A unique catalytic mechanism for UDP-galactopyranose mutase. *Nat Struct Mol Biol* **11:** 539-543

Sousa S, Cerqueira N, Fernandes P, Ramos M (2010) Virtual screening in Drug Design and Development. *Comb Chem High Throughput Screen* **13:** 442-453

Stahl M, Rarey M (2001) Detailed analysis of scoring functions for virtual screening. *J Med Chem* **44:** 1035-1042

Takayanagi T, Kimura A, Chiba S, Ajisaka K (1994) Novel structures of N-linked high-mannose type oligosaccharides containing alpha-D-galactofuranosyl linkages in *Aspergillus niger* alpha-D-glucosidase. *Carbohydr Res* **256:** 149-158

Tangallapally R, Yendapally R, Lee R, Hevener K, Jones V, Lenaerts A, McNeil M, Wang Y, Franzblau S (2004) Synthesis and evaluation of nitrofuranylamides as novel antituberculosis agents. *J Med Chem* **47:** 5276-5283

Thoden J, Holden H (1998) Dramatic differences in the binding of UDP-galactose and UDP-glucose to UDP-galactose 4-epimerase from *Escherichia coli*. *Biochemistry* **37:** 11469-11477

Tondi D, Slomczynska U, Costi M, Watterson D, Ghelli S, Shoichet B (1999) Structure-based discovery and in-parallel optimization of novel competitive inhibitors of thymidylate synthase. *Chem Biol* **6:** 319-331

Tronrud D (2004) Introduction to macromolecular refinement. *Acta Crystallogr D Biol Crystallogr* **60:** 2156-2168

Vagin A, Teplyakov A (2000) An approach to multi-copy search in molecular replacement. *Acta Crystallogr D Biol Crystallogr* **56:** 1622-1624

Vagin A, Teplyakov A (2010) Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr* **66:** 22-25

Veerapen N, Yuan Y, Sanders D, Pinto B (2004) Synthesis of novel ammonium and selenonium ions and their evaluation as inhibitors of UDP-galactopyranose mutase. *Carbohydr Res* **339:** 2205-2217

Villafranca J, Howell E, Oatley S, Warren M, Kraut J (1986) A structure-function study of dihydrofolate-reductase by protein engineering. *Phil Trans R Soc Lond A* **317:** *405-413*

Villoutreix B, Eudes R, Miteva M (2009) Structure-based virtual ligand screening: recent success stories. *Comb Chem High Throughput Screen* **12:** 1000-1016

Warren G, Andrews C, Capelli A, Clarke B, LaLonde J, Lambert M, Lindvall M, Nevins N, Semus S, Senger S, Tedesco G, Wall I, Woolven J, Peishoff C, Head M (2006) A critical assessment of docking programs and scoring functions. *J Med Chem* **49:** 5912-5931

Weston A, Stern R, Lee R, Nassau P, Monsey D, Martin S, Scherman M, Besra G, Duncan K, McNeil M (1997) Biosynthetic origin of mycobacterial cell wall galactofuranosyl residues. *Tuber Lung Dis* **78:** 123-131

Wilton D, Willett P, Lawson K, Mullier G (2003) Comparison of ranking methods for virtual screening in lead-discovery programs. *J Chem Inf Comput Sci* **43:** 469-474

Winn M (2003) An overview of the CCP4 project in protein crystallography: an example of a collaborative project. *J Synchrotron Radiat* **10:** 23-25

Wlodawer A, Minor W, Dauter Z, Jaskolski M (2008) Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *FEBS J* **275:** 1-21

Xu H (2002) Retrospect and prospect of virtual screening in drug discovery. *Curr Top Med Chem* **2:** 1305-1320

Yao X, Bleile D, Yuan Y, Chao J, Sarathy K, Sanders D, Pinto B, O'Neill M (2009) Substrate directs enzyme dynamics by bridging distal sites: UDP-galactopyranose mutase. *Proteins* **74:** 972-979

Yuan Y, Bleile D, Wen X, Sanders D, Itoh K, Liu H, Pinto B (2008) Investigation of binding of UDP-Galf and UDP-[3-F]Galf to UDP-galactopyranose mutase by STD-NMR spectroscopy, molecular dynamics, and CORCEMA-ST calculations. *J Am Chem Soc* **130:** 3157-3168

Yuan Y, Wen X, Sanders D, Pinto B (2005) Exploring the mechanism of binding of UDP-galactopyranose to UDP-galactopyranose mutase by STD-NMR spectroscopy and molecular modeling. *Biochemistry* **44:** 14080-14089

Yue Q, Kass I, Sampson N, Vrielink A (1999) Crystal structure determination of cholesterol oxidase from Streptomyces and structural characterization of key active site mutants. *Biochemistry* **38:** 4277-4286

Zhang Q, Liu H (2000) Studies of UDP-galactopyranose mutase from Escherichia coli: An unusual role of reduced FAD in its catalysis. *J Am Chem Soc* **122:** 9065-9070

Zhang Q, Liu H (2001) Mechanistic investigation of UDP-galactopyranose mutase from Escherichia coli using 2- and 3-fluorinated UDP-galactofuranose as probes. *J Am Chem Soc* **123:** 6756-6766

Zhang Y (2005) The magic bullets and tuberculosis drug targets. *Annu Rev Pharmacol Toxicol* **45:** 529-564